PREFACE
TO THE FOURTH EDITION

CONSIDERABLE development is taking place to-day in the upper reaches of the short-wave spectrum—the region of the micro-waves. Much of the information is not yet generally available, but the time seems opportune for the inclusion of some discussion of the technique of velocity modulated oscillators and the important subject of wave guides. Two additional chapters have therefore been incorporated to deal with these aspects of the subject.

PREFACE

THE study of short waves is a very fascinating one, dealing as it does with regions which, a few years ago, were considered unsuitable for radio communication. Gradually the queer and apparently inexplicable behaviour of these waves was reduced to a decently ordered basis. Further barriers to progress appeared. The frequencies in use became so high that the transit time of the electrons in the valves assumed a major importance. Yet methods have been evolved to surmount the obstacles and progress continues.

This book is an attempt to present a non-mathematical résumé of the position. A sound knowledge of ordinary radio technique is essential to a proper understanding of short-wave phenomena and this has been assumed to a great extent. Even so, many of the more specialized branches of the subject have had to be treated somewhat briefly, though the references quoted will enable the student to follow the matter up if he so desires. For proofs of many of the statements attention is directed to the author's *Modern Radio Communication* (Pitman) to which the present work forms, in large measure, a companion volume.

J. H. REYNER

BOREHAM WOOD
FEBRUARY, 1937

CONTENTS

CHAPTER V

CHAPTER VI

CHAPTER VII

CHAPTER VIII

CHAPTER IX

GLOSSARY

A COMPLETE definition of all the terms used throughout this book is impracticable and it has been assumed that the reader has a good working knowledge of ordinary theory. Some of the more specialized terms, however, are briefly elucidated in the following glossary. **The explanations given must not be taken as rigid definitions.**

A.C. Alternating current, i.e. a current which flows first in one direction and then in the other, usually changing in a rhythmic manner.

Amplitude. A term indicating the size or extent of an effect. If an oscillation is doubled in amplitude it means that the current and voltage are both doubled.

Attenuation. Reduction in amplitude.

Bridge. An electrical network comprising essentially two separate paths for the current. At some point along each path the potentials will be the same, and these points, therefore, may be connected together. Usually one of the arms of the bridge is variable or exhibits a varying effect with frequency so that the balance condition only applies in certain stipulated circumstances.

Capacitance. (Symbol: C.) The property exhibited by certain electrical elements of holding an electrical charge. Capacitance is usually deliberate as with a condenser—an arrangement of parallel plates in which, under the influence of an applied voltage, electrons will accumulate on one set of plates and remain there until they are allowed to return to their normal position by discharging the condenser. The greater the charge produced by a given applied voltage, the greater is the capacitance of the arrangement. *Stray capacitance* is the effect exhibited by two conductors at differing potential, such as the turns of a coil, the wiring of a set, etc.

Charge. An accumulation or deficiency of electrons. The former produces a negative charge, the latter a positive charge.

Choke. An inductance (see later) specifically designed to present a high impedance to some particular frequency or group of frequencies.

Conductance. (Symbol: G or g.) The susceptibility of a conductor to the passage of current. The better the conductance the

greater the current which will flow for a given applied voltage. Conductance is the reciprocal of impedance.

Coupling. Energy may be transferred from one circuit to another at radio frequencies by invisible means, e.g. by placing two coils together so that the magnetic field produced by one influences the other. Such circuits are said to be *coupled.*

D.C. Direct current, i.e. a current which flows steadily in one direction only.

Damping. An oscillation started in a circuit having no resistance would build up to an infinite value. In a practical circuit the oscillation is limited by the resistance and loss in the circuit. The total effect of these restraining influences is called damping.

Dielectric. The insulation between conductors. The capacitance of a condenser depends on the nature of the insulation between its plates. It is least with air and is always greater if any other material is used. The ratio of the capacitance with a solid (or liquid) dielectric to that with air alone is called the *dielectric constant* of the material.

Distortion. Speech or any other intelligence is communicated by series of oscillations following one another in a specific sequence. If this sequence is altered, or if any of the component oscillations are disproportionately amplified or transmitted, the desired effect is not correctly reproduced. This is known as distortion.

Dynamic Resistance. The effective impedance to A.C. of a coil with a condenser in parallel varies with the freqency. At a particular value, known as the *resonant frequency*, the inductive and capacitive effects are in opposition, and if the circuit had no losses, no current would pass. In practice the cancellation is not complete, and the circuit behaves as a high resistance. usually of several hundred thousand ohms, this being known as the *dynamic resistance.*

Earth's Field. The earth behaves as a huge magnet with magnetic poles near the rotational N. and S. poles. The magnetic field in the vicinity of the earth arising from this effect is called the earth's field.

Eddy Currents. The current produced by the movement of electrons in a conductor, a sheet of metal or an electrified layer under the influence of an external electric or magnetic field.

Emission. The giving off of electrons from a heated filament or cathode in vacuum. These electrons are drawn off in a valve by surrounding the cathode with an anode maintained at a suitable positive voltage.

Great Circle. A circle passing round the circumference of the earth. The great circle distance between any two points is the distance measured along such a circle passing through the said

two points on the surface of the globe. This is the shortest distance between the two points.

Henry. (Abbreviation: H.) The practical unit of inductance.

Hexode. A valve having six electrodes, usually a cathode, four grids and an anode.

High Frequency Resistance. Due to the magnetic field in and around a wire carrying an alternating current, eddy currents are induced in the wire itself, as the result of which the current does not flow uniformly through the wire but is more intense at the outer edge or *skin*. This *skin effect* becomes more pronounced as the frequency is increased. Because of this, the resistance of a wire, particularly when wound in a coil, is several times greater at high frequencies than when the wire is carrying d.c.

I. F. Intermediate Frequency. See "Superheterodyne."

Inductance. (Symbol: L.) A measure of the magnetic effect produced by passing a current round a coil of wire. The inductance is proportional to the diameter and shape of the coil and also to the square of the number of turns.

Inertia. Resistance to change. Just as a heavy weight is hard to set in motion owing to its mechanical inertia, so circuits containing inductance exhibit an electrical inertia necessitating electrical forces to set up a current through the coil or cause any change in the current already flowing.

Impedance. (Symbol: Z.) A property analogous to resistance but taking account of the effect of inductance or capacitance in the circuit on an alternating or oscillating current. (See "Reactance.")

Lag. Due to its inductive effect, the current produced in a coil by an alternating voltage does not pass through its various changes at the same time as the voltage, but lags behind. In a pure inductance the lag is one-quarter of a complete oscillation which is 90 electrical degrees, the complete oscillation being considered to occupy 360 electrical degrees.

Lead. Because of the finite time which must elapse in charging a condenser, the voltage developed across it (which is proportional to the charge) lags behind the current, or in other words, the current leads on the voltage. With a condenser having no resistance this lead is 90 electrical degrees.

Leakance. A term denoting the amount of leakage across the circuit, being the reciprocal of the insulation resistance.

Linear Law. When one effect is directly proportional to another it is said to obey a linear law, since a graph of the effect would be a straight line.

Load. Any circuit absorbing power in a valve or feeder is

known as a load. It may be a resistance, or a dynamic impedance of some type.

Magnetic Field. The magnetic influence surrounding a magnet or a coil carrying a current.

Megohm. (Abbreviation: MΩ.) One million ohms—the unit of high resistance.

Mercator's Projection. A representation of the world on a flat plane. In order to achieve this the scale has to become larger and larger towards the Poles, being infinitely large at these points since what should be represented as a point has become lengthened into a line. The shapes of the different countries are therefore distorted and in particular the shortest distance between two points is not shown on such a map by joining them together with a straight line. Actually, the great circle distance on a Mercator's projection appears as a sine wave.

Microfarad. (Abbreviation: μF.) The practical unit of capacitance, and one millionth of a farad (F). One micro-micro-farad ($\mu\mu$F.) is one millionth of a microfarad.

Microhenry. (Abbreviation: μH.) One millionth of a henry. The unit adopted for the small inductances used in tuned circuits.

Morse. Signals transmitted according to the Morse code used for telegraphy.

Mutual Inductance. An effect dependent upon the magnetic influence of one coil, known as the *primary*, on another, known as the *secondary*, due to the coupling between them.

Normal. A line at right angles to a given surface.

Optimum. Most satisfactory.

Parallel. When an electrical current has two alternative paths through a network, the paths are said to be in parallel.

Pentode. A valve having five electrodes, usually a cathode, three grids, and an anode.

Permeability. (Symbol: μ.) Certain substances respond to magnetic influences much more readily than air, and the magnetic field produced by a given current will be considerably greater with such substances. The ratio of the magnetic field produced with a given substance to that produced under the same conditions with air is known as its permeability.

Phase. The lag or lead between current and voltage is known as the *phase difference* between them.

Q. The symbol for the amplification or gain of a resonant circuit. $= L\omega/R$. (See "Resonance.")

Quadrature. Any two effects having a phase difference of 90° are said to be in quadrature.

Reactance. The effect on the current of an inductance or the capacitance is proportional not only to the value thereof but also to the frequency. The combined effect of the two is known as the reactance of the circuit. For an inductance it is $L\omega$ and for a capacitance it is $1/C\omega$. The combined effect of reactance and resistance constitutes the impedance of the circuit.

Resonance. The effects of inductance and capacitance on a circuit are exactly opposite, one taking a lagging current and the other a leading current. If the reactances of the two portions of the circuit are made equal their effects thus cancel out, and the current in the circuit is limited purely by the resistance. This is called resonance and clearly occurs at one frequency only—the *resonant frequency*. Due to this resonant effect the voltage developed across either the coil or the condenser is many times that induced in the circuit, and this is known as the *magnification* or *gain* of the circuit, usually represented by the symbol Q.

Series. Two circuit elements so arranged that the current has to pass first through one and then through the other are said to be in series.

Square Law. An arrangement in which the output is proportional to the square of the input is said to obey a square law.

Stranded Wire. An arrangement sometimes used to minimize skin effect by building up a conductor with a series of strands all insulated from one another and so twisted that each one in turn comes to the outside of the conductor.

Superheterodyne. An arrangement whereby the incoming high frequency currents are mixed with a local oscillation to produce a resultant oscillation of lower frequency which is then passed through an intermediate frequency amplifier tuned to this frequency. This i.f. amplifier can thus be fixed in its tune, since any incoming frequency can be converted to the required intermediate frequency by suitable choice of oscillator frequency.

Tangent. A line drawn from a point external to a curve which just touches the said curve.

Tetrode. A valve having four electrodes—usually cathode, two grids and an anode.

Time Constant. In a network including a condenser and resistance the charge or discharge of the condenser is slowed down by the presence of the resistance, and the time taken for the condenser to charge (or discharge) to approximately two thirds of the full amount is known as the time constant of the circuit, $t = CR$. A similar effect occurs with an inductance in which the current does not immediately rise to its full value owing to the electrical inertia. Here again, the time constant is that taken

for the current to rise to approximately two-thirds of its full amount, and in this case it equals L/R.

Transformer. An arrangement of coils coupled magnetically together so that by passing an alternating current through one an alternating voltage is developed in the second. At low frequencies the coils are both mounted on an iron core and the arrangement is much used for increasing or decreasing the voltage from the normal supply value around 200 to a low voltage for heating valve cathodes or to a high voltage for application to their anodes. At radio frequencies an iron core is not customarily employed, at any rate in short wave practice, and the transformers are used mainly for altering the effective impedance of a circuit for matching purposes.

Triode. A valve having three electrodes, a cathode, a grid and an anode. The anode attracts electrons from the cathode and the amount of the current flowing is varied by applying suitable potentials to the grid.

Ultraudion. Early types of American valve were known as *audions*, and a special form of oscillating circuit in which the tuned circuit is connected between anode and grid was called an *ultraudion circuit*.

Vacuo-Junction. A device for measuring small currents, consisting of a small heater wire in association with a thermo-couple —a junction of two dissimilar metals which develops a voltage in proportion to the heat. Since the heating is proportional to the square of the current the arrangement can be used for measuring current, and by making the heater wire very fine and enclosing it in a vacuum to minimize any radiation, alternating currents of as little as one milliamp can be measured.

ω. The symbol for $2\pi \times$ frequency—a factor which enters largely into a.c. theory.

SHORT-WAVE RADIO

CHAPTER I

WHAT ARE SHORT WAVES?

WITHIN the past fifty years man has learned to harness one of nature's most fascinating phenomena—the electromagnetic wave. From this the whole science of radio communication has grown up and, in fact, by far the greatest use of these waves to-day is for the communication of intelligence of some sort from one place to another, although comparatively recently other applications such as medical usage are beginning to appear.

It is as well to understand at the outset something of the nature of these electric waves. Any exhaustive discussion is apt to become intensively mathematical, but fortunately it is possible to obtain, without the aid of any great amount of mathematics, a reasonable mental picture of the processes involved.

The most convenient starting point for the explanation of many electrical phenomena is the *electron*, which may be considered as the ultimate unit of electricity, a tiny entity of which the exact nature is still a matter of conjecture. It is satisfactory for many purposes, however, to regard it as a minute particle far smaller than can be seen with a powerful microscope, and carrying with it its own tiny electrical charge.

All matter, whether solid, liquid or gas, is composed of these electrical particles in association with somewhat similar but heavier particles known as *protons*. The atoms of matter (at one time thought to be the smallest possible particle) themselves consist of a solar system comprising a proton around which are revolving a number of electrons in similar fashion to the revolution of the planets round the sun.

Since these electrons carry electrical charges, the whole miniature system exhibits an electrical force which causes the atom to set itself in a certain equilibrium with the adjacent

atoms. According to the number, rotation, and general disposition of the various electrons in an atom, so the forces developed differ slightly and we obtain different elements having distinctive physical properties. Obviously, certain groupings of the electrons will show a marked preference for associating themselves with other groupings, and conversely, will only link up with other types under great provocation, which shows why certain elements readily combine whereas others do not.

Electric Current.

Normally all the electrons in a mass of material are occupied with their own business of rotating around their respective protons and the material is in a general state of equilibrium. If, however, some of the electrons are caused to deviate from their normal paths, we obtain a drift of electrons which constitutes what we call an *electric current*. In metals and good conductors, the application of a suitable force will cause some electrons to leave their parent atoms altogether and wander through the material under the influence of the force, which is called the *electromotive force* or e.m.f.

In non-conductors or *dielectrics*, as they are called, the electrons will not leave their parent atoms except under extreme provocation, but they will displace themselves in their orbits, so that all the electrons in the material give a sort of lurch to one side. This momentarily causes a current known as a *displacement current*. If the e.m.f. is steady, then it will cause a momentary displacement current when it is first applied, but thereafter nothing will happen. On the other hand, if the e.m.f. is alternating, i.e. flowing first in one direction and then in the other, it will produce a similarly alternating displacement current, and in fact the material will behave to all intents and purposes as if it were a conductor. This possibility of obtaining current in a material which is apparently an insulator is important in radio practice, as will be seen later on.

Lines of Force.

One can continue to discuss the electron theory of matter indefinitely, but enough has been said to provide a starting

point, and we shall consider specific phenomena from time to time as occasion arises. For the present we are concerned with obtaining a mental picture of an electric wave based on this simple conception of electrons.

Now it has been explained that an electron carries with it an electric charge, in consequence of which it is exerting an electrical force in all directions. This force will repel another electron in the vicinity, but will attract a positively charged particle, such as a proton. It is customary to visualize electric *fields* as composed of a number of *lines of force*, running in the direction of the force. The stronger the field, the more lines there will be in a given space.

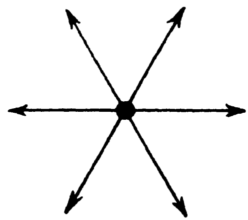

FIG. 1. CONCEPTION OF LINES OF FORCE

In the simple case of an electron, we have a number of such lines radiating from the electron in all directions, as shown in Fig. 1. Near the electron the lines are

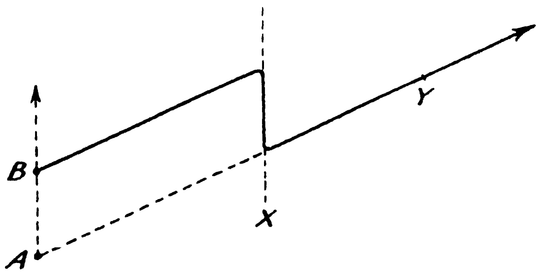

FIG. 2. ILLUSTRATING KINK IN LINE OF FORCE PRODUCED BY SUDDEN CHANGE

close together, indicating a strong field, while as we go farther away the field becomes progressively weaker.

Now let us consider an electron situated at point A (Fig. 2). Lines of force radiate in all directions, and we will just consider one of them passing as shown through X and Y and out into space. Now what happens if we move the electron suddenly to a point B? The lines of force accompanying the

electron must, of course, move with it, but it is impossible for this to happen immediately.

The whole system, in fact, has inertia, and it takes a little time for the straggly ends of the lines of force to get into position. Any electrical disturbance travels with a certain constant speed—the speed of light, which is 186 000 miles per second. Therefore a short fraction of time after the electron has moved from *A* to *B*, the information regarding this change will only have had time to reach the point *X*, and the point *Y* still farther out will be unaware that any change has taken place. Consequently, at *Y* the line of force is still where it was originally. At *X* and at all points nearer than *X* we know that the electron has moved so that the line of force is now in its new position, and at *X* there is a sharp changeover from the old to the new condition.

The important point to note is that this changeover is travelling outwards with the speed of light, and this is, in fact, our electric wave. There are many physical phenomena somewhat similar in character. Perhaps the most obvious is that of a wave at the seaside—the miniature wall of water which travels forwards at a more or less steady velocity. Still another analogy may be obtained with a rope lying on the ground. If the near end is given a sharp jerk a sharp kink will be produced in the rope, and this ripple will travel rapidly down the rope to the far end, and if the near end is jerked up and down at a rapid rate a series of waves will be generated, following one another down the rope.

This indeed gives a very good idea of an electric wave. Instead of just moving the electron from one point to another and leaving it there, we cause it to oscillate backwards and forwards, and this causes a ripple in the lines of force which travel out into space with the speed of light.

There is obviously one such ripple for each complete oscillation or cycle of the electron, and as long as the electron continues to oscillate, there will be a steady succession of waves radiating from it.

Wavelength.

With this conception in view, the idea of wavelength becomes comparatively simple. During the movement of the

electron through one complete cycle, the electric force in the wave itself will first be upwards (when the electron itself is moving upwards) and then downwards (when the electron is returning). The force is a maximum when the electron is moving very fast in the middle of its travel. It is obviously nothing when the electron has momentarily stopped at the top of its travel and is about to return. Then it reaches a maximum when the electron is travelling downwards at its highest speed, falls to zero again at the end when the electron

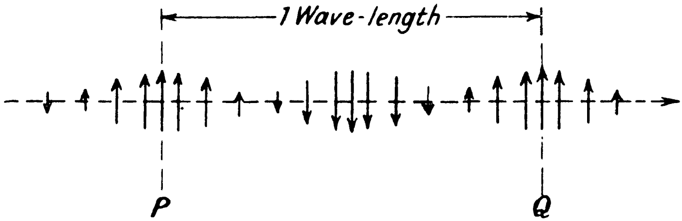

FIG. 3. ILLUSTRATING MEANING OF WAVELENGTH

is just turning round to come back, and then proceeds to grow to a maximum again, reaching the maximum point at the middle of the upward travel.

This is one complete cycle of operation, and since the information is being transmitted outwards with a definite velocity, we shall obtain at any point *P* on the route a succession of electric fields, firstly in one direction and then in the other, there being a gradual transition from one to the other. This is represented diagrammatically in Fig. 3, in which the length of the lines is proportional to the strength of the field and the arrows indicate the direction.

At the end of one complete cycle the point *P* will again be experiencing the maximum upward field. The original field, however, has by this time reached the point *Q*, and in between *P* and *Q* we therefore have a complete cycle of field strength. We call the distance between *P* and *Q* the *wavelength* of the wave. It is, as will be seen, the distance between the maximum field strength in any one direction and the next point of maximum field strength in the same direction.

It will be clear that the more rapidly the oscillations of the

electrons follow one another, the closer together will be the successive bands of field strength, because the wave is travelling at a constant speed. If the oscillations are slow, the first wave has had time to travel quite a long way before the second is generated, while if the oscillations are very rapid, the distance between the successive waves is obviously much shorter. In fact, wavelength and frequency are directly related to one another by the expression

$$f \times \lambda = c$$

where f = frequency of oscillation (cycles per sec.)
λ = wavelength of generated wave (metres)
c = velocity of wave = 3×10^8 metres per sec.

Transmission of Electric Waves.

It is obvious that a short wave is one having a short wavelength. The question is—how short? What sort of wavelengths are used in practice?

Actually, these range from a few centimetres to several miles, and the choice of wavelength is determined very largely by the distance to be covered. In fact, the transmission of these wireless waves is an important factor which we must consider briefly before going any farther.

In a practical transmitting system we do not have just one electron, but many millions of electrons, all caused to oscillate in unison under the influence of suitable external forces, and constituting a current as already explained. This current is caused to flow up and down a suitable aerial system as described later, and the aggregate effect of the kinks or ripples in the individual lines of force radiated from each electron add up to produce a powerful electric wave which will proceed to travel outwards from the transmitting point in all directions.

Now, it is clear that some of the radiation will be transmitted into the Earth where it will be rapidly absorbed by losses. Some will be transmitted directly upwards, where it will be wasted, and some will travel outwards along the Earth's surface. The Earth, however, is spherical, and if this radiation is proceeding in a straight line, it will very soon leave the surface of the Earth altogether and go out into space. Obviously, therefore, unless some other effect is coming into play, it is impossible to transmit more than a few hundred miles.

Yet, in practice, transmissions are carried out from one side of the world to the other, and the reason for this is that the wireless waves are reflected from the upper atmosphere, which is more or less permanently electrified. We shall discuss the mechanism of propagation in more detail in the next chapter, but it will suffice for the present if we assume that there is an electrical mirror at a height of some fifty miles above the Earth's surface, which reflects the waves so that they travel down again to the Earth, and it is because of this we are able to cover such long distances.

Now with long wavelengths of the order of several thousand metres, this reflection is quite good and consistent. Wavelengths up to 15 000 metres are in regular use and even longer wavelengths have been employed. In the neighbourhood of 300 metres, however, the reflection from the upper atmosphere becomes disturbed. Wavelengths of this order are no longer adequately reflected, but commence to penetrate into the electrified layer of gas, and in doing so are very seriously reduced in strength due to losses, as explained in Chapter II. At one time it was thought that no wavelength lower than about 200 metres could ever be used for transmission.

There is, however, a second electrified layer at a still greater height, and we find that as the wavelength is reduced below the critical value, the absorption begins to fall off, and the wave is transmitted through the first electrified layer without any appreciable loss. It travels on to the second layer, where it is again reflected, comes down to earth through the first layer, and arrives as a surprisingly good signal. In fact, we shall see that the characteristics of this class of transmission are quite different from those on the longer waves, and it is possible to cover very long ranges with relatively little power.

Actually, the transmission gets better and better as the wavelength is reduced, but there is naturally a barrier to an indefinite improvement. This occurs at wavelengths around 10 metres, which thus constitutes a second critical point.

Reflection is caused by an actual bending of the wave where it enters the electrified layer. The direction of travel is slowly but steadily changed until the wave is once more pointing towards the ground. As the wavelength is reduced, this bending becomes less and less, until at wavelengths of about 10

metres the deflection is no longer sufficient to return the wave to the Earth, and it is, in consequence, lost altogether.

Ultra-short Waves.

Ten metres, however, does not represent the limit of wavelength which we can generate, and there is thus a third range of wavelengths below 10 metres which are known as *ultra-short*. These have certain applications, although their range is very limited. In the absence of any reflection from the upper atmosphere, they can only be effective for the small distance before they leave the Earth's surface. This is a little over the visual range.

Even here, however, it seems that unsuspected forces are at work, and evidence has recently been put forward to show that reflection of ultra-short waves is possible. The subject is at present being investigated by various experimenters, and the results are not yet really understood, but with present technique the reliable range is something quite small.

In some ways this is an advantage, for we shall see later there is considerable interference on ordinary short waves, and a transmission of purely local character unaffected by any reflection from the upper atmosphere is of value for certain particular requirements.

Micro-waves.

These ultra-short waves extend right down to wavelengths of a few centimetres only. At frequencies corresponding to a wavelength of about 1 metre, ordinary valves become impracticable, and special forms of valve have to be used. This has led to a further classification, wavelengths below 1 metre usually being termed *micro-waves*. One advantage of such waves is that, owing to the short wavelength, very effective reflectors can be constructed without difficulty and the radiation can thus be concentrated in a very intense beam.

Such a system, operating on 17·4 cm., is in use between Lympne and St. Inglevert, a distance of 35 miles across the English Channel.

Classification of Waves.

As the knowledge of these micro-waves grows we are finding radically new techniques for their generation and

transmission. They are indeed becoming less like radio waves as we know them, and more like the light waves or the infra-red rays of the physicist. It is not improbable that, in the near future, they will become of such importance as to warrant a literature of their own.

Such then is, very briefly, our knowledge of the wavelengths suitable for radio communication, and the table herewith summarizes the classifications just enumerated.

Type of Wave	Wavelength (Metres)	Frequency*
Long waves	450 to 15 000	20 to 665 kc/s.
1st critical band	150 to 450	665 to 2 000 ,,
Short waves	12 to 150	2 000 to 25 000 ,,
2nd critical band	8 to 12	25 to 37·5 Mc/s.
Ultra-short waves	1 to 8	37·5 to 300 ,,
Micro-waves	Under 1 metre	Over 300 ,,

* kc. stands for *kilocycles*. 1 kc. = 1 000 cycles.
Mc. stands for *megacycles*. 1 Mc. = 1 000 000 cycles.

Strictly speaking, a frequency should be referred to as so many kc. or Mc. *per second*, and the abbreviations kc/s. and Mc/s. are commonly used. In actual discussion, however, the radio engineer usually omits the words "per second" and this convenient, if academically inaccurate, practice is occasionally adopted in this book.

It is significant that the long waves which were for a long time the only ones in practical use and which, even to-day, are far more familiar to the average radio engineer, occupy less than one-tenth of the available spectrum. The succeeding chapters will deal more precisely with the fascinating problems encountered in the remaining nine-tenths.

CHAPTER II

THE PROPAGATION OF WIRELESS WAVES

WE have seen in the preceding chapter how wireless waves are produced. and discussed briefly the mechanism of their transmission from one point to another. A complete understanding of short-wave transmission and reception requires a more detailed survey of this subject, so that the behaviour of this class of electric wave may be reasonably predicted.

The production of the electric wave itself is discussed both qualitatively and quantitatively in the next chapter, where the devices necessary for moving electrons backwards and forwards at the necessary rate are discussed. It will be clear that the greater the distance over which the electron can be caused to move, the larger will be the ripple in the line of force, and hence the more intense the wireless wave generated. Devices enabling the electrons to move through a relatively great distance are called *aerials*, and their particular application to short-wave transmission and reception is one of the most fascinating aspects of the subject.

For the present we shall confine our attention to the transmission of the waves after they have been generated. One of the fundamental laws regarding the radiation of waves from an aerial is that the strength of the wave falls off inversely as the distance from the transmitter, an effect which arises mainly from the fact that the waves spread out in all directions. In practice, the attenuation, or falling off, of the waves is greater than this due to losses of various kinds. The Earth, for example, is not perfectly conducting and absorbs energy from the "feet" of the waves as they travel over its surface. The upper atmosphere, in turn, also absorbs its proportion of the energy, so that the ideal state of affairs is never attained. One of the advantages of short-wave transmission, however, is that the field strength at the far end is nearer to this theoretical value than in the case of longer waves, for reasons which will become evident

The importance of the electrified layers in the upper atmosphere has already been mentioned. Were it not for these layers, the waves would never be reflected back to the Earth again and long distance transmission would be an impossibility. Let us examine this electrified upper atmosphere in more detail.

Troposphere and Stratosphere.

At the surface of the ground and for a height of five or six miles, the atmosphere is more or less homogeneous. The density and pressure gradually decrease and the temperature also falls, normally approximately one degree for every 300 ft. rise. The presence of wind and air currents generally, however, serves to maintain the various gases of which the atmosphere is composed in a mixed state, so that the general nature of the atmosphere remains much the same. This region immediately next to the earth is called the *troposphere*, and its upper edge marks the limit of cloud formation.

Above this limit we have a region in which the air is still, and therefore tends to separate itself out into its various constituent gases in order of density. For this reason the upper region is called the *stratosphere,* and it has certain properties which make it of particular interest to the radio engineer. There is no temperature gradient, but there is still a gradual decrease in pressure with increasing height, while the layer formation just mentioned means that the upper reaches of this region, which extends for several hundred miles above the Earth's surface, are of a light and easily ionized character.

Ionization is the name given to the action which takes place when some of the electrons constituting the atoms of a gas detach themselves and roam about at large. At normal temperatures and pressures, ionization is small, and in any case in the troposphere it is of small importance, because of the continual presence of air currents which would quickly dispel any ionized patches. As the pressure decreases, however, the electrical forces between the atoms are reduced. Indeed, the pressure itself is a manifestation of these forces, and a point is reached at which the hold of the parent nuclei on their planetary electrons is so feeble that they can readily be detached by external influences.

Radiation from the Sun is one of the most powerful of these agents, this being partly due to direct electron bombardment and partly in the form of ultra-violet radiation. Ultra-violet waves, being very short, are rapidly absorbed in their transit through the Earth's atmosphere and only a very small proportion ever reaches the ground, but up in the higher reaches of the stratosphere this radiation is present in large quantities and effectively ionizes the gas. Moreover, when a gas has been ionized, the random electrons themselves, in their transit to and fro, encounter other atoms of gas, which they, in turn, ionize by collision. At each collision the electron is slowed down and ultimately it is reclaimed by a wandering ion.

In an ionized gas this ionization and re-combination is continually taking place, and an equilibrium state is attained depending upon the power of the ionizing influence.

The Ionosphere.

Now we have evidence to show that at a considerable distance above the Earth's surface there is a layer of gas which is more or less permanently ionized. This is known as the *ionosphere*, and actually is itself constituted in a series of layers. The lowest layer appears at around 100 km. and is quite heavily ionized, due, it is thought, to direct corpuscular bombardment from the sun. That is to say, clouds of electrons or similar bodies actually radiated from the sun in showers penetrate the outer atmosphere and produce this layer of ionized gas known as the *E layer*.

At a considerably higher distance—some 200 km. up—is another layer of electrified gas known as the *F layer*, which is ionized by ultra-violet radiation from the sun. This radiation does not penetrate sufficiently far to maintain appreciable ionization lower down, and thus we have two quite distinct layers. The existence of the second layer was not suspected for some time even after the presence of the first had been considered fairly well proven. It was largely the researches of E. V. Appleton which demonstrated the existence of the two distinct layers, and it was he who gave them the names of E and F layers by which they are now generally designated.

Now there is an important difference between the action of the two layers. In the first place, the lower layer, occurring in

the region where the gas is denser (so that the molecules are closer together), is subject to much more rapid re-combination.

In consequence of this more rapid re-combination, the ionization is very dependent on the ionizing influence—in this case the Sun. As the Sun's rays strike the atmosphere, the gas very quickly attains the ionized condition, and when sunset occurs the gas equally quickly restores itself to its normal un-electrified and therefore insulating condition.

The F layer, on the other hand, being situated in a region of very much lower gas pressure, is less dependent on the withdrawal of the ionizing influence. Naturally, of course, some time after sunset the re-combination is appreciable and the atmosphere will gradually return to its insulating state, but this does not occur until several hours afterwards, so that there is a very considerable time-lag. It should be noted that this time-lag does not occur at the beginning of the period. The re-ionization follows very quickly on the heels of sunrise.

How Bending is Produced.

Let us consider what happens when an electric wave passes into the E or F layer. The ionization is, of course, gradual, and the waves in their travel will gradually begin to encounter a number of free electrons. These electrons they will set in motion in sympathy with the electric forces existing. Moreover, since there must of necessity be some inertia in the system, the currents set up will lag behind the voltage in the wave by which they are produced, and we can split the current into a component in phase, which will absorb energy and attenuate the wave, and a component lagging by 90°.

Now in any dielectric subject to electrical stress, there is a *displacement* current caused by the movement of electrons which do not actually leave their parent atoms, but merely change their orbits. This displacement or *capacitance* current leads by 90° on the e.m.f. Hence the lagging component of the free electron current set up in an ionized layer of gas, such as the Heaviside layer,* will operate to reduce the effective displacement current, which is clearly the same as if the dielectric constant of the medium were reduced.

* The existence of the ionosphere was first postulated by Heaviside, and it is often referred to as the *Heaviside Layer*.

The velocity of an electromagnetic wave is given by the expression

$$v = 1/\sqrt{(\mu\kappa)}$$

where μ = permeability

κ = dielectric constant of the medium.

Consequently, a wave entering the ionized belt starts to move with greater velocity, because of the reduced dielectric constant. Moreover, the farther it penetrates, the faster it

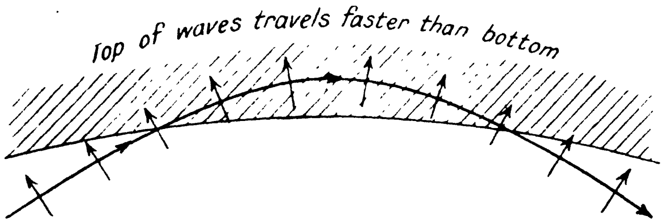

FIG. 4. SIMPLE CONCEPTION OF BENDING IN IONOSPHERE

gets, the net result being that the tops of the wave travel faster than the bottom, so that it bends round as shown in Fig. 4, and ultimately leaves the layer again in a downward direction. The amount of bending produced, and also the energy absorbed in the process, depends on the wavelength.

Attenuation

Where the frequency of the wave is low, the electrons in the ionized gas undergo a number of collisions every cycle, and consequently appreciable energy absorption takes place. At the same time there is a fairly sharp reflection. The wave is very rapidly bent round and hardly penetrates the layer at all before emerging in a downward direction.

At much higher frequencies, however, the conditions are different, for there is time for the free electrons to make several complete oscillations under the influence of the electric wave before encountering an ion, and consequently the energy absorption is quite small. The reduction in dielectric constant is also smaller, so that the bending is not so marked, and the wave travels for a considerable distance before coming out again. In fact, as the frequency is progressively raised, a

point is reached where the bending is insufficient to return the ray to ground at all, and it continues through the layer in the manner shown in Fig. 5.

We are now able to consider what happens to the various classes of electric wave specified at the end of Chapter I.

Long Waves.

Wavelengths longer than 300–400 metres are characterized by the fact there is time for several collisions among the

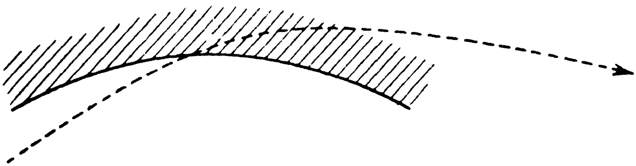

FIG. 5. IF THE BENDING IS INSUFFICIENT, THE WAVE DOES NOT RETURN

electrons in the E layer every cycle. Under these conditions, reflection is sharp, and during the time the wave is actually in the layer the attenuation is quite appreciable, but as it only penetrates to a small depth the loss is not serious.

With this class of wave, in fact, the attenuation is proportional to $\sqrt{f}$, and hence the longer the wavelength, the less the attenuation. This led to the use of increasingly longer wavelengths in the early days of long-distance transmission, and even to-day world-wide broadcast services are carried on on wavelengths between 10 000 and 20 000 metres. It is possible to transmit half-way round the world with such waves, the two principal disadvantages to their use being the very heavy atmospheric disturbances which are liable to be encountered, and the increasing power necessary. The field strength of a wireless wave is proportional to the frequency of the current. Hence, as we increase the wavelength, the field strength falls off rapidly and the current must be correspondingly increased.

Critical Region.

In the region 200–300 metres, a marked absorption occurs. This is because the mean free time of electron travel in the

E layer coincides roughly with the period of the oscillation in the electric wave itself. Consequently, a resonance effect results which produces relatively enormous losses. The exact resonant frequency depends upon the ionization conditions of the layer, and therefore varies from day to day or even from hour to hour. Wavelengths in the critical region therefore are reflected very erratically, and they are quite useless for reliable transmission.

Short Waves.

Below the critical belt, we reach the conditions where the free electrons have time to make several oscillations before colliding. Under these conditions the energy loss is small and becomes increasingly small as we increase the frequency. Such waves, as this, therefore, are reflected from the E layer with relatively little loss, the attenuation being proportional to $1/f^2$.

With short waves, therefore, the least attenuation is obtained with the shorter wavelengths. As we increase the frequency, we reach a point where the bending is insufficient to return the wave to earth at the E layer, and the wave therefore continues onwards to the F layer where the reflection is completed, and the wave ultimately returns earthwards. In fact, for the normal waves used for commercial practice, the E layer has little effect, most of the reflection occurring at the F layer.

Because of the questions of fading, polarization, and similar phenomena which we shall discuss later, it does not follow that the shortest wavelength is necessarily the best; but, other things being equal, the best results are obtained with the shorter wavelengths.

Ultra-short Waves.

This improvement continues until we reach the second critical period, where the bending, even in the F layer, is insufficient to return the wave to earth. The wave, therefore, travels outwards into space and is lost. The frequency at which this occurs corresponds to a wavelength of 10 metres. Once again the transition is gradual, and it is not safe to use wavelengths below 12 or 13 metres for long-distance work.

On the other hand, occasions have been recorded where transmission has been achieved on wavelengths below 10 metres, and it is not till we reach 7 or 8 metres that real ultra-short wave technique begins to apply.

There is evidence to-day of some form of reflection or refraction even with these ultra-short waves, but the results are by no means so reliable or on such a large scale as with the normal short waves. We shall discuss the matter further in Chapter IX.

Observation of Reflection Conditions.

Detailed research has been carried out for many years on these reflecting layers. The F layer has been shown to be twofold, one region, known as the F_1 layer being about 200 km. high and the F_2 ranging from this to 600 km. in height. There is also evidence of reflections from the troposphere, the changes in velocity here arising from atmospheric rather than electronic variations. Such layers have been termed the C and D layers.

Their detection arises from improved technique. The method is to generate a short pulse of waves and to observe, on the screen of a cathode ray tube, the time elapsing between the emission of the pulse and the return of its reflection. With low level layers, the reflection time is only a few microseconds, involving very rapid sweep of the spot across the cathode ray screen.

The existence of these additional layers, however, does not affect the basic operation of reflection, and we need not consider it further here. The reader who wishes to do so should refer to the papers below.*

Skip Distance.

It will be clear that there is a period between the transmitter and the receiver over which no signals can be received. Close to the transmitter the ordinary ground wave will provide the necessary signal strength. At long distances the signal

* "Measurements of the Equivalent Height of the Atmospheric Ionized Layer," Appleton, *Proc. Roy. Soc.*, March, 1930. "Measuring the Reflecting Regions in the Troposphere," Friend and Colwell, *Proc. I.R.E.*, December, 1937. "Tropospheric Reflections," Friend and Colwell, *Proc. I.R.E.*, October, 1939.

will arrive by reflection from the upper layers, but there is a gap between the point at which the ground wave dies out and that at which the first reflected wave begins to appear.

Radiation from the transmitter may be sent out at all angles, and obviously the high angle radiation will come down to earth first. Thereafter, the whole surface of the Earth will be covered by waves which have left at progressively smaller angles until the maximum range with a single reflection results with the wave which left tangential to the Earth's surface.

There is, however, a critical angle above which no reflection results, when the wave enters the Heaviside layer too sharply

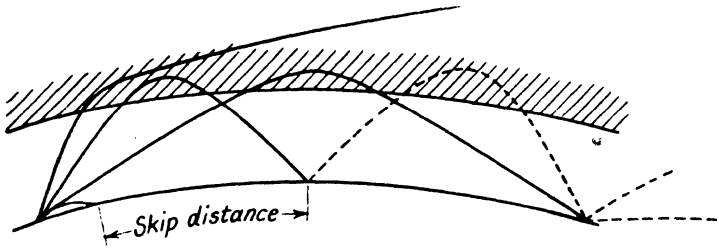

FIG. 6. ILLUSTRATING SKIP DISTANCE

to be completely reflected and is lost. Because of this there is the gap just mentioned between the end of the ground wave and the beginning of the sky wave, and this is usually known as the *skip distance*. It may range from a few hundred miles to several thousand, depending upon the wavelength and the atmospheric conditions.

It will also be clear from the explanations just given that wireless waves can continue to travel round the Earth by successive reflections. A wave which has been reflected from the Heaviside layer will travel down to earth, and here again will be reflected by a very similar process. It will then journey out into space once more, and again be reflected by the upper atmosphere. It is possible for this process to continue a large number of times, or alternatively for a wave to arrive at the receiving point from two different directions, one having come by a single reflection and the other by several high angle reflections, as shown in Fig. 6.

Such reception may lead to distortion, as we shall see; and, apart from the wastefulness of the process, there is a tendency to-day to eliminate high-angle radiation by the use of directional transmitting aerials, which concentrate the radiation in a direction between 3° and 15° from the horizontal, since practice indicates that the best direction lies within these limits.

Echoes.

One of the forms of distortion which arises from the multiple paths of the ray through the Heaviside layer is what is known as a *quick* or *multiple echo*. Since the time of transit from the transmitting to the receiving point is clearly slightly different according to whether the wave has travelled with one reflection only or with a number of successive reflections, there will in effect be a number of signals at the receiving point following closely on the heels of one another. The ray which has suffered the least reflection will obviously be the strongest, since the opportunities for attenuation in transit have been less, but there will under normal conditions be a number of other weaker waves following immediately afterwards.

In ordinary Morse transmission this causes no serious difficulty. It produces, in fact, a slight lengthening of the signal which is not serious enough to give appreciable distortion. Increasing use is being made, however, in long distance communication of facsimile transmission, in which sheets of typescript or similar matter are scanned by passing a point of light across the image in a series of lines so that the information is converted by a photo-cell into signals of varying duration and spacing, which, when re-assembled in the correct sequence at the receiving end, reproduce the original script.

It is clear that any lengthening of the signal in such a transmission will result in a blurring of the characters, which may be quite serious. The remedy, as already mentioned, is to reduce the high angle radiation at the transmitter, so that the received wave is composed as far as possible of waves which have only suffered a small number of reflections.

In telephone transmissions the effect of a quick echo is to produce a hollowness in the speech, as in the case of an audible echo of very short duration. Similar hollow tone is often deliberately introduced in broadcasting to produce eerie effects.

One-seventh Second Echoes.

A different type of echo is that which is obtained from a signal which has gone more than once round the world. It travels along to the receiver, where it is duly recorded, and then carries on until it makes a complete circuit of the Earth and arrives at the receiver again, producing an echo $\frac{1}{7}$-second later. Because of this precise delay, arising from the product of the velocity of the wave and the Earth's circumference, these echoes are often referred to as *one-seventh second* echoes.

This form of echo may be weak or strong, according to the conditions. It is at its worst when the transmitter is not very far away, for then under favourable conditions of the upper atmosphere the echo signal may be transmitted right round the world with little attenuation, so that it arrives at the receiver nearly as strong as when it passed the first time. Several such echoes may be obtained at successive $\frac{1}{7}$-second intervals. The favourable conditions required are that the great circle line between the stations coincides roughly with the shadow band (i.e. the dividing line between light and dark, as it is explained later) and, with such conditions, wave-lengths between 15 and 18 metres show round-the-world echoes quite markedly.

There is little or no remedy for this class of echo. Directional methods are of no avail, for the signal arrives on its second time round from the same direction as it did on its first transit from the transmitter.

These round-the-world echoes must not be confused with back echoes due to the reception of a signal which has travelled the opposite way round the Earth and arrived at the receiver from the rear. The time interval elapsing here depends upon the relative distances in the forward and reverse directions, and if the transmitter and receiver are practically at opposite sides of the Earth, then the two waves travelling in opposite directions will arrive practically at the same time. Indeed, it is sometimes the practice to reverse the beam at the transmitting and receiving ends at certain times of day, and deliberately receive the opposite way round.

The use of beam reception, of course, entirely eliminates the back echo, and since practically all commercial services

embody some form of directional reception, this class of echo is not of serious moment.

Reference should be made finally to very long delay echoes sometimes experienced at intervals of between 10 and 30 sec. after the original signal. The reason for these is not understood, and they are so long after the original signal that they cause comparatively little trouble, being, in fact, a species of atmospheric disturbance. They are sometimes called *Störmer* echoes, after the engineer who first discovered them.

Scattering.

The phenomenon of reflection from the ionized upper atmosphere is sometimes regarded as being due to re-radiation. The free electrons are set in motion by the wireless waves and abstract the energy from it. They, in turn, then become small transmitters and will re-radiate waves of practically the same strength as before, minus a small inevitable loss.

The majority of the electrons will move in the plane of the wave, so that the greater part of the re-radiation will be as if the wave had continued on its path, but on a fresh course. There will, however, be a certain proportion of the electrons which will vibrate in other directions, and these will radiate waves in more or less random fashion. The strength of such radiation is small, and in regions where the normal wave is being received, their presence is practically unnoticed. Within the skip distance, however, where no reception is being obtained by normal channels, this scattered radiation will provide a very weak and somewhat erratic signal. The effect is somewhat similar to the diffusion of light using a matt reflector. The main beam will still continue, but there will be a great deal of scattered radiation in the vicinity of the reflector.

Polarization.

We now come to an important factor in the propagation of wireless waves, namely, the plane of polarization. The ripple or kink in the line of force produced by the motion of the electrons is obviously in the same direction as the motion of

the electrons themselves. That is to say, if we have a vertical aerial and we send the currents up and down in the oscillating manner already described, we shall produce electric waves in which the fields are vertical. Such a wave is said to be vertically polarized, and if we want to receive that wave we must use a similar vertical aerial at the receiving point.

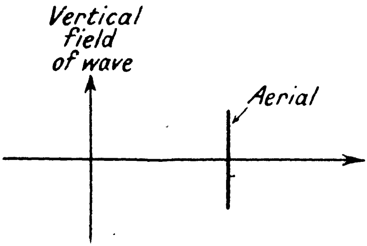

FIG. 7. A VERTICAL AERIAL RESPONDS TO VERTICALLY POLARIZED WAVES

Action of Electric Field.

The meaning of the term electric field has not been explained, since it is assumed that the reader will be familiar with basic radio theory. In brief, we may summarize the position as follows. If a battery is connected across a piece of wire, the e.m.f. of the battery will cause a movement of the electrons in the wire which constitutes an electric current. An electric field is an abnormal condition or state of strain tending to produce a similar motion of any electrons which come within its influence, and if such a field is moving past a wire, an e.m.f. will be induced into the wire, which will set up a current proportional to the strength of the field in similar fashion to that produced by a battery.

It is necessary that the wire shall lie wholly or mainly in the direction of the field. Thus, if a vertical field is moving horizontally, as illustrated diagrammatically in Fig. 7, then the maximum current is induced when the wire is also vertical. If it is inclined at an angle, then the current induced is less, being, in fact, proportional to the effective height, namely, the vertical distance between the top and bottom ends of the wire. Clearly, therefore, if the wire is horizontal, no voltage at all will be induced.

It must be understood that this explanation is very much simplified, and the reader who has not a clear understanding of the action should refer to suitable textbooks. The matter is discussed at greater length in *Modern Radio Communication*, Vol. I, by the author (published by Pitman).

Circular Polarization.

Now, it is found that in the process of reflection from the upper atmosphere, the plane of polarization of the wave is affected. Moreover, it is not usually a simple twist, such as, for example, a wave entering the layer vertically polarized and emerging horizontally polarized. The action is somewhat more complicated, for it is found that the plane of polarization is usually rotating. Thus, at any given receiving point one belt of electric field may be vertical, but the next wave will be slightly rotated, and so on until after a certain period the wave is horizontally polarized. The plane of polarization continues to alter until after another lapse of time the wave is again vertically polarized, although now in the opposite direction from what it was originally, and so on.

A wave of this nature is said to be *circularly polarized,* and most of the waves encountered after reflection from the upper atmosphere are of this form, or of a modified form known as *elliptically polarized,* in which the actual strength of the wave changes according to the direction of polarization. For example, it may be that the field strength when the wave is vertically polarized is greater than when it is horizontally polarized, so that not only have we a rotating plane of polarization, but also a continuous cyclic variation in amplitude.

Effect of Earth's Field.

Actually, this rotation of plane of polarization is produced by the effect of the Earth's magnetic field. Hitherto, we have assumed the motion of the electrons in the ionosphere to be directly dependent on the forces contained within the wireless wave itself. The motion, however, will obviously be also controlled to some extent by the presence of the Earth's magnetic field, and it has been shown that a wave entering an ionized medium in the presence of a magnetic field will split into two components, each having different paths through the medium and suffering different attenuations.

If the wave is travelling transversely across the magnetic field, the effect is merely to split it into two components, each of which is still plane polarized. On the other hand, if the wave is travelling with the field, two circularly polarized waves are produced, rotating in opposite directions.

Since in practice the wireless waves comply with neither of these two conditions, but enter at various angles, the net result is a very complex one, resulting in an elliptically polarized wave, i.e. a wave of which the plane of polarization is rotating, and also the strength is varying according to the position in the cycle.

Fading.

Now we have seen that if a horizontal wire is held in the path of the vertically polarized wave, no signal will result. Similarly, if an ordinary vertical aerial is acted upon by a horizontally polarized wave, the result will again be zero. Hence it will be clear that when we are dealing with a circularly polarized wave there will be a continuous change in the strength of the induced signal.

Quite apart from any variation in the strength of the wave, the signal will be a maximum when the wave is vertically polarized (assuming a vertical receiving aerial) and it will gradually decrease in strength until it is zero, when the wave is horizontally polarized, after which it will commence to increase again. If, in addition, the wave is elliptically polarized, the variation in strength will be either accentuated or minimized, according to the nature of the cyclic variation in strength.

It is this effect which gives rise to the fading which is commonly observed with short-wave signals and also on medium wavelengths of the order of 400 to 1 000 metres. Beyond this limit the effect is not by any means so marked as one might expect, since the upper atmosphere produces a very sharp reflection with such waves and has therefore little time to influence the plane of polarization.

The fading experienced may be either slow or rapid, according to the conditions in the upper atmosphere. Very rapid fading produces a sort of gurgling noise, which completely distorts and may even obliterate the signals, while very slow fading does not as a rule produce any serious distortion, but only a change in strength, culminating in a complete fade-out which may last for some minutes, and has been known to last for days!

The intermediate type of fade occurring every 10 or 20 sec.

is quite common, and this is accompanied by distortion because it is found that the exact effect on the transmission is very dependent on frequency. As we shall see later, the communication of intelligence necessitates the transmission of a carrier wave and a number of side frequencies, the strength and displacement of which provide the necessary information to be communicated.

It is possible, however, for the carrier wave to be affected when the side wavelengths are not, and vice versa, which results in very marked and quite unpleasant distortion. This action is known as *selective fading*, and is, of course, particularly obnoxious when receiving speech or music. It is, however, possible to overcome it by the use of special aerial receiving arrangements, some of which are discussed later in Chapter V.

Daily and Seasonal Variation.

We may conclude the chapter with a brief summary of the way in which the performance of wireless waves is predicted. We have seen that the shorter the wavelength, the less the bending in the upper atmosphere, and we have also seen that the most satisfactory results are obtained with the least number of reflections, since this gives the least attenuation and absence of quick echoes. Clearly, therefore, the best wave to use is the shortest wavelength which will just give the required reflection.

Unfortunately, the conditions are not the same for daylight and darkness, and the belts of light and dark are differently distributed according to the season of the year, even at the same latitude. Hence it is rarely possible to find a single wavelength which will maintain adequate communication throughout the twenty-four hours.

We can summarize the conditions as follows—

ALL DAYLIGHT. By daylight is meant the strong intense light round about middle day in this country. It extends for a greater period on each side of noon as one goes nearer the equator. The maximum distance which can ever be in the all-daylight zone is approximately 6 000 miles, and for this coverage a wavelength of 14–15 metres would be used. Shorter waves than this would have a skip distance exceeding the

6 000 miles and in any case would be getting dangerously near the second critical zone, where reflection from the ionosphere is incomplete.

The attenuation during daylight is greater than at any other time, but by the use of waves which produce the minimum number of reflections the losses are kept as low as possible.

TWILIGHT ZONE. This zone includes early morning, late afternoon, and the early hours of darkness. The attenuation is much less on all wavelengths in this zone, and particularly so for waves below 20 metres. The twilight zone extends over much greater distance than the daylight zone, and once again we use the shortest convenient wave. For maximum range—which may embrace the full 12 000 miles of the Earth's semicircumference—we use waves between 15 and 18 metres, while for shorter distances longer waves up to about 40 metres are employed.

As already explained, if two stations are situated so that the path of the wave is in the twilight zone all the time, the transmission conditions are exceedingly good and $\frac{1}{7}$-second echoes are very liable to be formed.

DARKNESS ZONE. As the darkness becomes more intense, the ionization in the upper atmosphere becomes less and hence the bending of the wave is not so great. The shorter waves are therefore barred from our use—not from any consideration of attenuation or distortion, but simply because they will never come back. This leads to a gradual increase in the wavelength used as darkness falls, and within the ordinary darkness zone, wavelengths of 20–60 metres are customary. As the darkness becomes very intense, it is necessary to use still higher wavelengths, and, in fact, the critical wavelength below which the bending is insufficient may be as high as 30 metres.

Tremellen Charts.

Special charts are prepared for analysing the possibilities of short-wave transmission. The world is represented on the customary Mercator's projection on which the great circle lines which constitute the shortest distances between various points are in the form of a sine wave, as shown in Fig. 8. We

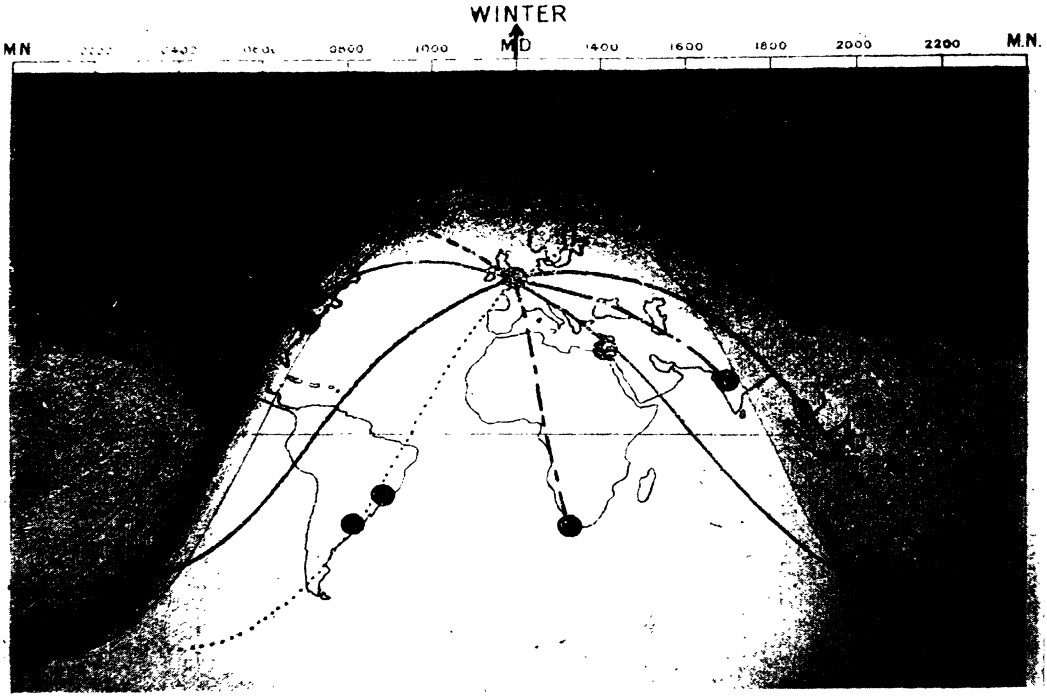

Fig. 8. Tremellen Chart
(Marconi's Wireless Telegraph Co.)

can prepare transparencies to lay over this chart, on which are shown shaded areas, the density of which represents the condition of the atmosphere, indicating a gradual transition from full daylight, through twilight and darkness to full darkness.

By plotting first of all the great circle distance on the map, we can get an idea of the path which the wave will have to follow. Then by laying on top of the chart the appropriate Tremellen light and shade chart, we can see whether the path will be in daylight, in darkness, or partly in both. Data has been collected showing what course of action to adopt in the latter case, and it is possible by the use of these charts to determine the best wavelength for the particular conditions obtained.

It is not practicable to discuss these charts more fully here, and the reader who wishes more concise information should refer to "Studies in Radio Transmission" by T. L. Eckersley, *Journal I.E.E.*, Vol. 71, Sept., 1932, p. 405.

CHAPTER III

AERIALS AND FEEDERS

WE must now consider the practical methods adopted for the production of wireless waves. We showed in Chapter I that such a wave was produced by suddenly changing the position (or uniform motion) of an electron. In practice, we achieve this by causing the electrons to oscillate backwards and forwards some millions of times per second.

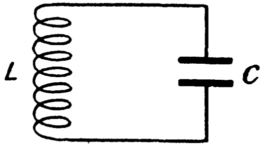

FIG. 9. SIMPLE OSCILLATORY CIRCUIT

In achieving this very rapid oscillation, use is made of the well-known property of *resonance*, with which the reader will be familiar. Fig. 9 illustrates a simple oscillatory circuit, comprising a condenser and inductance connected in series to form a closed circuit. If the condenser is assumed to acquire a momentary charge, it will proceed to discharge through the inductance.

Oscillatory Discharge.

A current flowing through an inductance produces a magnetic field, and since this field has to be built up, the inductance presents an appreciable inertia to the growth of current, which therefore does not suddenly acquire its full value, but takes some fraction of time in the process.

The same electrical inertia prevents the magnetic field from collapsing immediately, and therefore prolongs the current after the condenser has been discharged completely, with the result that the over-shoot carries the charge into the opposite plate of the condenser and leaves it charged in the reverse direction. The discharge process then repeats itself in the opposite direction with similar results, so that a continuous succession of surges backwards and forwards occurs in the circuit. The time period of the surge depends upon the values

of inductance and capacitance in the circuit, the frequency being given by the expression

$$f = \frac{1}{2\pi\sqrt{(LC)}} \text{ (neglecting the effect of resistance)}$$

where f = frequency in cycles per sec.
L = inductance in henries.
C = capacitance in farads.

Sometimes the oscillation we require is actually generated by permitting a circuit to oscillate in this manner. The current will obviously slowly decrease in value at each oscillation due to the inevitable loss caused by the resistance in the circuit, and this loss must be made up by suitable means, some of which we shall discuss later in Chapter VI. On other occasions we have available a rapidly oscillating voltage from some source—as, for example, in the receiving aerial where the passing wireless waves induce such a voltage—and we are concerned to make the current produced by the voltage as large as possible.

Once again we tune the circuit and obtain maximum current in accordance with the ordinary laws of a.c. theory, the current in a circuit containing inductance, capacitance, and resistance being

$$I = \frac{V}{R + j[L\omega - (1/C\omega)]}.$$

The current is clearly a maximum if we can reduce the reactive term to zero by making $L\omega$ equal to $1/C\omega$, which arrives at the same condition as before, viz., $\omega = 2\pi f = 1/\sqrt{(LC)}$.

For more detailed treatment of resonance and a.c. theory as applied to radio practice, the reader is referred to *Modern Radio Communication*.

The oscillating circuit therefore provides us with a means of causing the electrons to move rapidly backwards and forwards, and so to generate the electric waves which we require. Each electron carries its own line of force, and consequently the more electrons we can provide (i.e. the larger the current) the greater will be the field strength produced, since if all the electrons are moving together their individual fields will

produce an additive effect. Equally important is the question of the distance through which the electrons can be moved, for obviously if the oscillating circuit is of small dimensions, so that the travel of the electrons is limited to a few inches, the disturbance produced will not be very marked.

We must arrange, therefore, to move the electrons through as large a distance as possible, and this is done by increasing the scale of the construction to a suitable degree. A simple condenser consists of two parallel plates situated relatively close together and enclosing an area of a few square inches. As we increase the distance between the plates, the capacitance decreases, but we can bring it back to its original value by a corresponding increase in the area of the plates. Continuing this process we arrive at a condenser having plates, of which the area is measured in square feet, the distance between them also being measured in feet. Clearly, if we have made the capacitance of the condenser the same as before, the constants of the circuit are in no way affected, but the electrons in their travel in and out of the condenser plates are now forced to move over a distance of many feet, and the wireless wave generated will be appreciable.

Use of Aerials.

Such an arrangement is called an *aerial*. For long-wave transmission the upper plate is replaced by a network, or even a single wire, at a considerable height, while the lower end is connected to Earth. On such wavelengths, aerials of very considerable height are used. The Rugby world-wide long-wave radio transmitter, for example, uses 800 ft. masts with a network of wire at the top to form the upper plate. As we reduce the wavelength, however, certain difficulties begin to creep in which modify the whole construction.

So far we have assumed that the capacitance is concentrated at the end of the wire and the necessary tuning inductance is also concentrated in the form of a coil, usually located at the bottom end. Actually, any length of wire has an inductance, and when we commence to erect elevated structures such as aerials, the inductance of the connecting wires themselves become quite appreciable. In similar fashion there is a capacitance to earth from every part of the wire and not only from

the upper portion. In addition, there is a capacitance between one part of the wire and remoter parts of the same wire.

In view of this, and bearing in mind that as the frequency is raised the inductances and capacitances required become smaller and smaller, it will be understood that a single straight wire alone may constitute an oscillating circuit. The wire itself will possess self-inductance, while there will be capacitance between various portions of the wire, particularly between the end portions. Such an arrangement, therefore, has the requisite features for an oscillating circuit and, if voltage is induced therein by any convenient means, oscillating current will flow.

Current in an Aerial.

Where is this oscillating current going? There is no complete circuit, and the idea of a current in an apparently isolated piece of wire is perhaps a little difficult to understand. Remembering the capacitance effect between the ends of the wire, however, the operation becomes more readily visualized. When the capacitance is charged, there is an excess of electrons at one end of the wire, the atoms at the other end of the wire being deficient in electrons and therefore exhibiting a positive charge.

When this charge begins to dissipate itself, it will do so by a flow of electrons from the negatively charged end of the wire to the positive end, and because of the inductance in the wire this current will gradually build up to a maximum and will then continue beyond the equilibrium position until the opposite end of the wire has become negatively charged, reversing the conditions in the customary manner. A further discharge will then take place, restoring affairs to their original condition, and so current will rush up and down the wire until the energy has been entirely dissipated, unless there is some external source of energy which is maintaining the oscillation.

There is more to be learned from this simple wire, however, for the capacitance and inductance are uniformly distributed along the length of the wire and are not concentrated at any point. Therefore, the charged condition of the wire is not due to an accumulation of electrons at the extreme end. Actually, the unnatural or strained condition starts from the centre of

the wire and becomes progressively greater as we reach the end. Near the centre only a few of the atoms are deficient in electrons, while at a corresponding distance the other side of the centre point there is an equivalent surplus. As one progresses farther out, so the surplus or deficiency becomes greater, until at the extreme ends of the wire we have the maximum accumulation. The distribution of electric charge over the wire is, in fact, gradually changing.

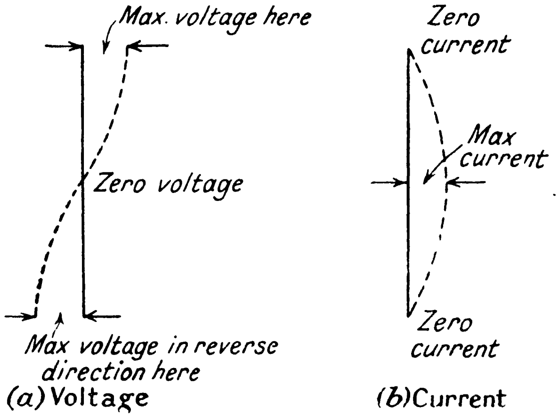

FIG. 10. NON-UNIFORM DISTRIBUTION OF VOLTAGE AND CURRENT ON WIRE

Now when a condenser is charged there is an electric force tending to restore the charge to its normal condition. A condenser can only be charged by the application of a suitable e.m.f., and develops an equal and opposite e.m.f., tending to restore the atoms to their normal equilibrium condition. This e.m.f. is proportional to the charge, and hence there will be a voltage on the wire which is zero in the centre, and will gradually increase as we progress along the wire. The voltage at the two ends will be equal and opposite, and there will thus be a distribution of voltage as shown in Fig. 10 (*a*).

In similar fashion, when the wire discharges the current will not be uniform. The centre of the wire will have to carry all the electrons which have passed from every part of the wire.

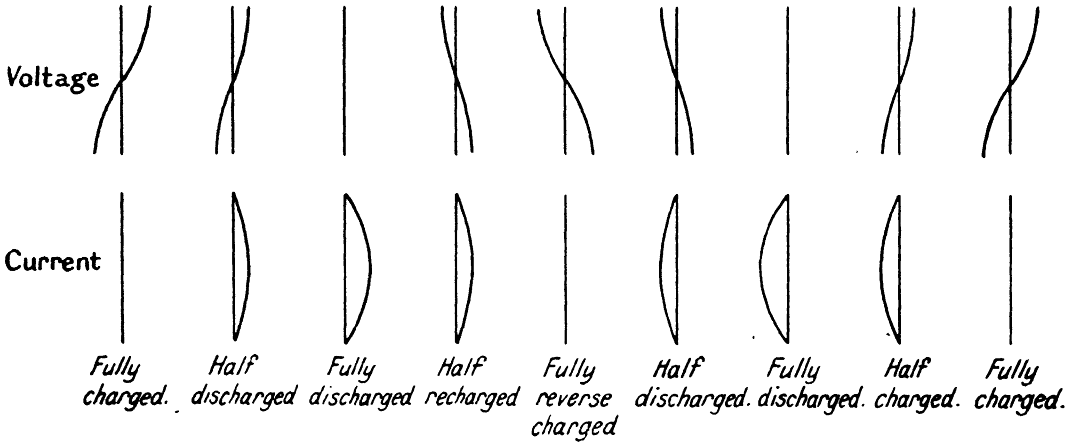

FIG. 11. VOLTAGE AND CURRENT AT SUCCESSIVE INTERVALS ON WIRE OSCILLATING NATURALLY

Portions of the wire farther out will only have to carry the electrons which have passed them coming from the extreme portions, and the ends of the wire, of course, will carry no current at all. Consequently, the current in the centre of the wire will be a maximum, and it will gradually decrease to zero at the ends of the wire as shown in Fig. 10 (*b*).

Standing Waves.

It must be clearly understood that we are talking about a spacial distribution of current and voltage quite apart from any variation with time. For example, consider the discharge of a charged wire, various stages of which have been set out in Fig. 11. When the wire is fully charged there is no current anywhere along its length. As it commences to discharge, current will flow, and this current will be greatest in the centre of the wire and will be progressively less as we reach the ends. The discharge continues and the current increases. It does so *at every part of the wire* and therefore maintains the same relative distribution, being a maximum in the centre and zero at the ends. The voltage has now fallen at every part of the wire to a fraction of its original value.

Ultimately we arrive at the condition where the maximum discharge current is flowing. Here again the same distribution is obtained, while the voltage is, momentarily, zero all along the wire. The current, however, cannot immediately cease, but dies away gradually in well-known oscillatory fashion, charging the wire in the opposite direction as it does so, and this decrease continues until we finish with a fully reverse-charged condition and zero current. The current at each part of the wire has increased from zero to a maximum and decreased to zero again, but the actual value of the maximum becomes greater and greater as we approach the centre of the wire.

Thus both current and voltage in the wire change in a cyclic manner. The variation with time is sinusoidal, and so also is the spacial distribution, and, in fact, with a wire oscillating naturally, the current and voltage are given by the expressions

$$v = V \sin \omega t \sin (\pi d/l);$$

$$i = I \cos \omega t \cos (\pi d/l);$$

where l is the length of the wire;
d is the distance from the centre;
$\omega = 2\pi \times$ frequency;
$t =$ time from the commencement of the oscillation.

This spacial distribution of current and voltage is called a *standing wave*, the wavelength being the distance between points of the same polarity. With the simple, freely oscillating wire, we clearly have half a wavelength, for which reason such wires are often termed *half-wave* aerials.

Forced Oscillation.

Now suppose that instead of allowing the aerial to oscillate naturally, we induce current in it from some external source. If the frequency is the same, the current and voltage distribution will be unaltered, and we shall obtain a standing wave half a wavelength long, as before. If, however, the frequency is not the same, the standing wave produced will correspond to the altered frequency, and the particular length of wire we are using may accommodate more or less than half a wavelength.

For example, if the frequency is doubled, the wavelength will be halved, and our wire will now accommodate one complete wave, as shown in Fig. 12 (*a*). In fact, a wire oscillating under the influence of an external e.m.f. will set up standing waves along its length dependent merely on the frequency of the current. If the length of the wire happens to be an odd fraction of a wavelength, the distribution will arrange itself with the point of no-current at the end—a state of affairs which must obviously exist. Such a partial wave is shown in Fig. 12 (*b*).

Reactance of Wire.

The reactance of a wire under conditions of forced oscillation clearly depends on its length. If it is a multiple of half a wavelength long, there will be no current at the input end of the wire, so that the reactance will be infinity (as in the case of a parallel tuned circuit).

If the length is an odd multiple of a quarter wavelength, current will appear at the input end with no applied voltage,

so that the reactance is clearly zero. (Actually, of course, the inevitable presence of resistance will necessitate some applied voltage.) This corresponds to a series resonant circuit.

With any other length the wire exhibits a finite reactance depending on its length. Over the sections $n\lambda/2$ to $(n + \frac{1}{2})\lambda/2$ the reactance is inductive, while from $(n + \frac{1}{2})\lambda/2$ to $(n + 1)\lambda/2$ the reactance is capacitive.

Nodes.

The points of zero current (or voltage) are termed current (or voltage) *nodes*.* They clearly occur exactly one wave-

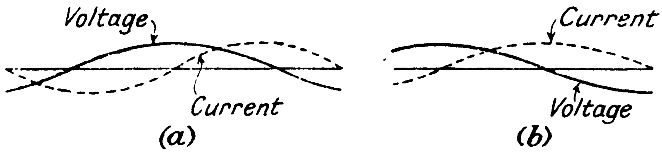

FIG. 12. STANDING WAVES ON WIRE UNDER FORCED OSCILLATION

length apart, this wavelength being related to the frequency of the current in accordance with the usual relation

$$\text{Wavelength (metres)} = \frac{3 \times 10^8}{\text{Frequency (cycles)}}.$$

Thus the wavelength of standing waves is the same as that of the travelling waves radiated by the wire.

The reader must distinguish between the condition of forced oscillation (when the waves produced depend only on the frequency of the current and not on the length of the wire) and free oscillation when the wire oscillates naturally. In the latter case the frequency automatically adjusts itself to provide one-half wave only, as shown in Fig. 11.

In practice, the length of a wire oscillating naturally is not exactly half a wavelength, but is slightly less than this owing to what is called *end effect*—a tendency to obtain a slightly increased capacitance at the extreme ends of the wire, so that the assumed uniform distribution of capacitance is no longer strictly correct. For a wire to radiate a given wavelength

* Which may be memorized, a little crudely, as "no darned current (or voltage)."

when oscillating naturally its length should be adjusted to about $0{\cdot}47\lambda$, where λ is the wavelength.

Practical Forms of Aerial.

This distribution of voltage and current on a wire becomes increasingly troublesome as we reduce the wavelength. It means, for example, that if we increase the length of the lead by even a few feet, we may appreciably alter the conditions, and this possibility has always to be borne in mind in short-wave operation. At medium and long waves the trouble does not arise, for the wavelengths in use are much longer than the length with which we are dealing. Even the height of 800 ft. quoted for the Rugby Station is small compared with the wavelength of some 15 000 metres which is used for that particular transmission.

The first point at which we come up against this problem is in the design of the transmitting aerial system itself. We desire to use a long wire so that the electrons may be caused to move through a considerable distance. Any appreciable length of wire, however, has in itself sufficient inductance and capacitance to cause the formation of these standing waves, and we reach an apparent impasse.

Consider the radiation from a one wavelength wire of the type shown in Fig. 12(*a*). Over the first half wavelength the current is all flowing in the same direction at any instant, and therefore the electric fields produced will all add up. In the second section of the wire, however, the current is flowing in the opposite direction, and therefore the waves produced will be in opposition to those set up by the first part of the wire. Consequently, except in an area quite close to the aerial, the net radiation from a wire such as this would be zero. It is clear, therefore, that with a simple aerial we are limited to a height of half a wavelength, a distance which may be as little as 20 or 30 ft.

Tiered Aerials.

The remedy is to use wires which are folded or bent back on themselves in such a manner that only the portions of the wire carrying current *in the same direction* are effective in producing radiation. Such aerials are known as *tiered aerials*,

and will be discussed further in the next chapter. A simple form is shown in Fig. 13. The first half wavelength of the wire is straight. The second half wavelength is bent round in a form of loop so that it comes back just on top of the end of the first half wavelength. The third half wavelength continues upwards in the same line as the first. The fourth is again bent round, and the fifth continues up, and so on.

Now the current in *AB*, *CD*, *EF*, etc., is all in the same direction, and therefore the radiation produced by these several portions will all add up and produce a wave of a strength several times greater than would be obtained by one section alone. The radiation from the looped portion cancels itself out. Considering the first portion, for example, the radiation from the wire *PQ* is cancelled out to a large extent by that from *RS*, and that from *ST* by that from *UV*. Cancellation is not complete because the current is not uniform over the loop, but it will be clear that the greater proportion of the unwanted reverse radiation is eliminated by this means. Radiation from the horizontal portions of the wires is also cancelled out, but in any case it would not be troublesome, because it is radiating directly upwards and downwards where it will do no harm.

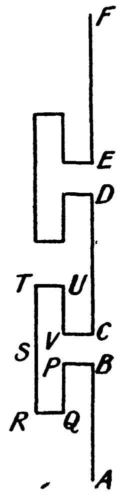

FIG. 13
TIERED AERIAL

High-angle Radiation.

This question of the radiation from the wire is a point which might conveniently be considered here. The main radiation from any wire goes out at right angles to the wire in all directions, but there is actually radiation in an upward and downward direction. Each individual electron radiates lines of force all round itself, and each one of these lines of force is subjected to a kink or ripple if the motion of the electron is suddenly changed. The belt of electric field produced is therefore vertical on the plane through the centre of the radiating system, and will be inclined backwards on a constant radius from the transmitter as we increase the angle

with the horizontal. The strength of the electric wave at an angle θ with the horizontal would actually be $\cos\theta$ times the strength in the horizontal plane, because the kink in the line of force will still be vertical and the effective wave front will be only $\cos\theta$ times as much. Consequently, at an angle of 90°, i.e. immediately above the radiating aerial, there will be no radiation at all.

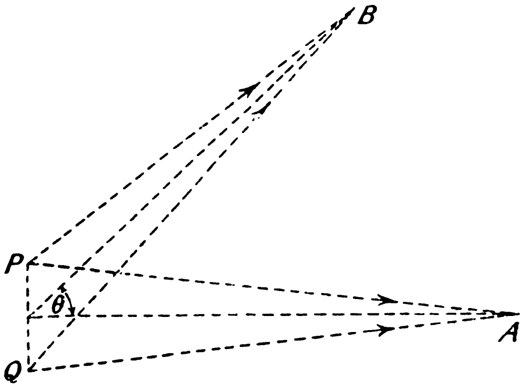

FIG. 14. ILLUSTRATING LOSS OF UPWARD RADIATION DUE TO PHASE DIFFERENCE

Now with short-wave radiation, when the length of the radiating aerial is comparable with the wavelength of the wave, further complications come in. With an ordinary aerial of relatively short length, the waves radiated by the individual electrons add up and produce a combined wave which again varies according to the cosine law just mentioned. With a short-wave aerial, however, the length has an important effect.

Take the case shown in Fig. 14, where we have a half-wave aerial, and consider the radiation from two portions nearly at its end. (At the end, of course, the current is zero.) At a point A in the horizontal plane the waves from P and Q have both travelled the same distance and will therefore add up in the normal manner. When we consider the radiation at an angle, however, to a point such as B, it is clear that the waves from the top of the aerial have to travel less distance than those from the bottom. In fact, the difference in distance is

PQ sin θ, as will be seen from Fig. 14, which means that the two waves will not arrive at the receiving point together, but one will be slightly behind the other.

Now when we add two sine waves of the same frequency, the resultant depends upon the phase angle between them. If they are rising and falling together, i.e. in phase, then the resultant is the sum of the two. If they happen to be exactly out of phase, so that when one is rising the other is falling, the resultant is clearly zero, and in between these two conditions the resultant will be somewhere between the maximum possible (the sum of the two) and zero.

It will be seen, therefore, that in an upward direction the effective radiation is less, due to the interference between the waves from the opposite ends of the aerial, so that quite apart from the cosine law already mentioned, there is a further restriction of the upward radiation.

Nevertheless, there is still a considerable amount of upward radiation left and, as we shall see in the next chapter, particular care is taken to minimize high angle radiation for long distance communication, so that the transmission shall be effected as far as possible by waves which have suffered only a few reflections.

Effect of the Earth.

So far, we have considered the radiation from an aerial in free space, but the presence of the Earth has considerable bearing on the results, due to the fact that the Earth is not by any means a perfect conductor.

On long wavelengths the effect is small, but with the increasing frequency of the currents used with short-wave working, Earth losses become troublesome. Not only is the resistivity increased, but the dielectric constant is high and may vary from 2 or 3 up to 40, according to the moisture content.

In consequence, the radiation from the aerial is very considerably affected and, in fact, approximates to the effect which would be produced by the aerial itself operating in conjunction with an image situated at the same distance *below* the level of the ground and carrying current in exact opposition.

It will be clear that one immediate effect of this is that radiation at the ground level is reduced practically to zero, since all radiation from the aerial proper is counteracted by radiation from the image; and this in some ways is an advantage, since, as already explained, it is desirable for long distance communication that the radiation shall be concentrated in a slightly upward direction.

The angle at which the radiation is most effective obviously depends upon the height of the aerial from the ground, since this determines the effective distance between the aerial and its image, and hence the phase relationships between the two radiations. In practice, empirical data is used in designing aerials, as the exact effect depends upon the wavelength employed, nature of the soil, and various other local conditions. For more detailed information on this subject, the reader cannot do better than refer to *Short-wave Wireless Communication*, by Ladner and Stoner (published by Chapman & Hall), an authoritative and very practical treatise.

It should be noted, of course, that this conception of an image aerial is one of convenience. Actually, the effects arise from the presence of currents induced in the Earth in the immediate vicinity of the aerial, and it is found that the effect of these currents can conveniently be estimated by assuming the existence of an image as described.

It follows, however, that the higher we make the aerial, the less will be the influence of the ground, and it is found that if the aerial is raised some three to three and a half wavelengths above the Earth, the discrepancies between theoretical radiation from a free aerial and the actual radiation obtained are quite small.

Aerial Resistance.

The resistance of an aerial differs from that of an equivalent closed circuit very considerably, by reason of its open character. This gives rise to the radiation of electromagnetic waves, as already explained, which absorbs energy from the system. In addition, these radiations induce currents in the stays, the Earth, and near-by objects, which accounts for a further loss of energy.

All these effects are proportional to the square of the current

and can, therefore, be considered as due to an additional resistance, the total effective resistance being the sum of the real and radiative resistances. Of this total, only that portion responsible for the radiation of electric waves is really useful, and this is usually termed the *radiation resistance.*

The calculation or measurement of effective resistance is a difficult matter. Many writers have investigated the problem, which is a very mathematical one, but beyond certain general rules no really accurate data have resulted. The effective resistance clearly depends on the current, and since this is different at various parts of the aerial, it is the custom to refer all calculations to the maximum r.m.s. current (i.e. at the current antinode) which occurs in the centre of a half-wave aerial or at the base of an earthed quarter-wave aerial.

On this basis the radiation resistance is some 36 ohms for a quarter-wave aerial and 72 ohms for a half-wave aerial, values far in excess of the conductor resistance. Hence the wire used for the aerial is relatively unimportant, being determined mainly by mechanical considerations. The loss resistance, however, is quite serious, being of the same order as the radiation resistance, and in practice the total effective resistance is about 50 ohms for a quarter-wave aerial and 150 ohms for a half-wave aerial.

For aerial arrays comprising several half-wave sections in series, the calculations are complicated by the interaction between the various sections, and the reader should refer to the bibliography at the end of this chapter for further information.

A quantity of interest is the *dynamic resistance* of an aerial considered as a tuned circuit, which requires to be known when matching feeders to the aerial. If L and C are the inductance and capacitance per unit length, then the aerial can be shown to be equivalent to a tuned circuit having a dynamic resistance $4L/CR$, where R is the total effective resistance.

For a vertical wire near the Earth, Howe has deduced the expressions

$$L = 2[\log_e(2l/d) - 1] \times 10^{-3}\mu\text{H.},$$

$$\text{and } C = \frac{10^{-5}}{18[\log_e(2l/d) - 1]}\,\mu\text{F.},$$

where l and d are the length and diameter of the wire in cm.

To match the feeder to the aerial, we connect it either across part of the aerial or between one end and earth. In the latter case we are effectively connecting the feeder across half the aerial, assuming a half-wave aerial, since the centre point of the aerial is also at zero potential. The resistance between these points is called the *base resistance*, and since the total dynamic resistance is $4L/CR$, the base resistance will be one-quarter of this value $= L/CR$.

If the feeder is tapped across a portion of aerial as shown in Fig. 18, the impedance of the tapped portion h may be taken as $(4L/CR)\,(4h^2/\lambda^2)$, which is based on the assumption that the arrangement is equivalent to a tuned circuit tapped $\dfrac{h}{\lambda/2}$ up.

High-frequency Feeders.

Up to now we have discussed the aerial as an isolated entity, but it is clear that we must have some method of supplying the necessary current. Moreover, if we are using any special form of aerial, it may be necessary to situate this an appreciable distance away from the transmitter itself, and under such conditions the length of the lead-in wires will certainly be comparable with, or even several times greater than, the wavelength being used.

It is necessary, therefore, to connect the transmitter to the aerial with correctly designed feeders. One such arrangement would be obvious from the discussion earlier in the chapter on the distribution of current and voltage in the wire. By continuing the aerial system by some exact multiple of a wavelength, we can leave ourselves in a similar position at the ends of the wire.

Fig. 15 represents a half-wave aerial with a current feeder. The current is introduced at the middle point of the aerial, and two feeder wires are made an exact multiple of half a wavelength. By this means the standing waves set up on the wires produce a current maximum at the input end, so that the conditions in the coupling coil are exactly the same as if the coil were actually in the aerial.

The standing waves on the two feeder wires will be in opposition to one another over the length of the feeder,

and hence the radiation from the feeder itself will be negligible.

An alternative method is to supply the energy at a voltage feed point (current node). The obvious way of doing this is

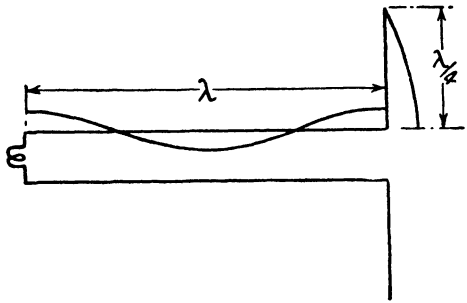

FIG. 15. TUNED FEEDER SUPPLYING A HALF-WAVE AERIAL

to tap the feeder (now $\lambda/4$ longer than before) across the ends of the aerial, but with such a system there is liable to be interaction between aerial and feeder or appreciable radiation from the feeder itself. There is, however, a slight modification which results in a practicable feeder, this being shown in Fig. 16.

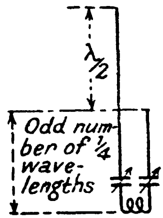

FIG. 16 ALTERNATIVE METHOD OF FEEDING A HALF-WAVE AERIAL

One of the feeder wires is connected to the end of the aerial, the length of the feeder being an odd multiple of a quarter wavelength, so that the point where it is connected to the aerial is a voltage maximum. The second feeder runs parallel with the first, and is the same length, but terminates in mid-air. Standing waves will occur on this wire in opposition to those on the live feeder, thereby cancelling the radiation and leaving only the aerial proper to be effective.

This latter form of aerial is often called a Zepp aerial, from the fact that it was first used on the *Graf Zeppelin*.

Terminations.

We have assumed that a current injection is required, and it is customary to terminate the feeder at the input end with

a coupling coil and two small series condensers as in Fig. 16. The coil is for inducing the current from the transmitter circuits, and the condensers for adjusting the tune of the feeder, two condensers being used for symmetry. The inductance and capacitance are chosen to resonate with the frequency in use, so that there would be a voltage across the coil and an equal and opposite voltage across the condensers, leaving zero voltage at the commencement of the feeder, which is exactly what we require. If the length of the feeder is not exact, so that our voltage node occurs a little way along the wire, this can be compensated by slight variation of the setting of the tuning condensers.

It is clearly possible to inject the voltage from the transmitter in other ways as, for example, by connecting across some point of high voltage. In such circumstances the feeder would have to be a quarter of the wavelength longer or shorter, so that the input end was a current node (i.e. a point of maximum voltage). Such arrangements, however, would usually give too tight a coupling between the aerial and the transmitter circuit, and the method first described is nearly always used, since it enables a variable coupling to be obtained, and thereby the best adjustment for each condition.

The spacing between the wires has little effect on the operation of the feeder, although there is an optimum spacing which gives the greatest efficiency. We shall, however, discuss this point later in the chapter.

Untuned Feeders.

The tuned type of feeder has been discussed first, since it is a natural development from the short-wave aerial itself. There are, however, several disadvantages with this form of connection, the first being the obvious one that the length must be accurately adjusted, the second being that since the feeder carries the full aerial current, the losses will be quite appreciable.

With any oscillating circuit the current is out of phase with the voltage by nearly 90°, and the power in the circuit is represented only by that component of the current which is in phase. Hence when supplying power from the transmitter to the aerial, it is really only necessary to provide energy

equivalent to this in-phase component. This can be done by using the properly terminated transmission line, sometimes called an *untuned feeder*. This consists of the same two wires as before, but they are connected to the circuit differently. How can a mere difference in connection make two wires behave in a totally different manner?

Let us consider what happens when we attempt to transmit a voltage along a pair of wires. The wires themselves contain inductance, while there is small capacitance between them.

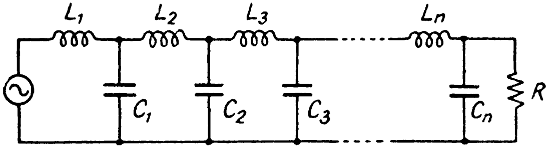

FIG. 17. ILLUSTRATING FEEDER AS A SERIES OF SECTIONS

These effects are not concentrated, but are uniformly distributed over the whole length, and for our purpose we can consider the feeder as broken up into a large number of small sections, each containing a small series inductance and shunt capacitance, as shown in Fig. 17.

Suppose that at the beginning of this feeder we suddenly apply a voltage. This voltage will send a current along the wire tending to charge the condenser C_1. The presence of the inductance, however, will delay the operation by a small fraction of a second, and the condenser also will take quite an appreciable time to charge, so that there will be a definite, though very small, time-lag in the transmission of the voltage over this section.

Meanwhile the condenser C_1 is proceeding to discharge into the second section L_2C_2 with a similar small but perceptible time-lag, and so the voltage is transmitted down the wire by successive handing on of the information from section to section until we finally reach the far end. At this point the local conditions determine what is to happen.

Suppose there is nothing connected across the far end. The last condenser C_n, having been charged completely, will at once commence to discharge by sending current back into the

preceding section, and this in turn into its predecessor, so that the travelling impulse, or wave, is reflected back along the line in the same way as a wave in a pool of water is reflected by a sharp boundary.

A similar reflection would occur if the line at the distant end were short-circuited. In this case the penultimate condenser would discharge through the inductance of the last section (the last condenser being short-circuited). The inductance would maintain the current after the condenser had been discharged in the usual manner, so that the condenser would charge in the opposite direction. It would then discharge and transmit the wave back along the feeder in the same manner as before, though now in opposite phase.

In practice, we usually terminate the feeder with some form of impedance, which will absorb energy from the line. If this impedance is high, the energy is not dissipated rapidly enough, and we still get a little reflection of the first type, while if it is low the energy is absorbed too rapidly and we get reflection of the second type. Clearly there is a critical value, when the energy is absorbed exactly at the rate at which it arrives, when we obtain no reflection at all.

This desirable state of affairs is obtained when the terminating impedance is made equal to the impedance of the line. At first sight this might appear somewhat indefinite, but if we remember that the line is composed of a number of sections of similar characteristics, it will be understood that the impedance of the line is independent of the length, and solely determined by the inductance and capacitance of the sections. This impedance is known as the *characteristic* or *surge impedance*, and is actually given by a very simple expression*

$$Z_0 = \sqrt{(L/C)}$$

where L and C are the inductance and capacitance of the line per unit length.

This impedance has the characteristic of a pure resistance, and hence, with reflectionless working, we have to terminate

* This expression assumes that the resistance of the feeder is negligible in comparison with $L\omega$ and that the leakance is also negligible in comparison with $1/C\omega$. Both these assumptions are justified in radio practice.

our line by a resistance (or an impedance of resistive character) equal to this critical value.

Both input and output to the line must be matched in this way, and if we do this we are able to transmit the voltage along the wire without reflection, irrespective of the length.

Travelling Waves.

The distribution of voltage and current along a transmission line requires to be clearly understood. If we apply an alternating e.m.f. to the input of a feeder, this voltage will be transmitted along the wire by the process just described. There is, however, a small time-lag, as we have seen, in the transmission, so that at any point in the wire the voltage will arrive slightly after it started.

At the distant end, therefore, the voltage will vary in strict accordance with the impressed e.m.f. at the sending end (assuming the feeder is correctly terminated), but a fraction of a second later, the voltage being transmitted along the feeder by a *travelling wave*. This wave will move with a velocity just a little less than that of wireless waves in free space. If the resistance of the feeder and the leakance between the wires is comparable with the reactances due to inductance and capacitance, this slowing down is appreciable, and in telephone work where we are dealing with relatively low frequencies the velocity is considerably reduced. In radio technique, however, the loss is small, and the velocity of propagation is practically the same as that of light.

One important fact arises from this finite velocity of travel. Although the length of the feeder is immaterial from the point of view of reflection, two points on a feeder will not exhibit the same voltage or current *at any given moment*. At any instant the voltage and current will be varying sinusoidally along the wire as in the case of a tuned feeder, but the wave is travelling so that sooner or later every part of the wire experiences the normal variation of voltage impressed at the sending end. With a tuned feeder this is only obtained at the voltage antinodes, because the wave is not travelling, but is stationary. At every other point the voltage variation is less than the full amount, and at the voltage nodes there is no voltage whatever *the whole time.*

If we have two feeders, therefore, both supplied with e.m.f. from the same source, the voltage and current at the far ends will only be in phase if the feeders are of exactly the same length or if one is an exact multiple of one wavelength longer or shorter than the other. This fact has to be remembered when dealing with aerial arrays, in which various parts of the system have all to be supplied with current in the same phase.

Relation Between Tuned and Untuned Feeders.

The tuned feeder, which we discussed earlier, is actually a particular case of the incorrectly terminated line. If the feeder is not connected to the correct impedance, the voltage and current are reflected at the far end, travel back to the input end, and are then reflected again, so that we obtain a series of successive waves surging back and forth along the wire. If we are applying the alternating voltage on the sending end, the waves transmitted will encounter reflected waves which have left a short time previously, and the result will be complex interferences which will be difficult to predict.

Obviously no substantial amount of energy will be transmitted to the receiving point, since most of the energy supplied will be frittered away in transit.

There is just one case where something useful happens, which is when the length of the line is a definite multiple of a quarter of a wavelength. In these circumstances the time taken by the reflected wave to travel back is such that various current and voltage nodes coincide, and instead of indeterminate interference we obtain standing waves, as already described.

Matching the Line.

The value of the characteristic impedance depends upon the type of line, being of the order of 600 ohms for open lines, and about 80 ohms for tubular feeders. Either of these values is much lower than the impedance of either a tuned circuit or an aerial, and hence we have to use line transformers to match the feeder to the equipment. The customary valve transmitter is provided with a tuned circuit in the anode of the output valve which has an impedance at resonance of

L/CR, which is resistive in character. By coupling to the coil a secondary winding giving an effective step-down ratio of n, the effective impedance across the secondary will be $1/n^2$ times the circuit impedance L/CR, and it will therefore be quite simple to choose the ratio so that the secondary impedance is equal to a characteristic impedance of the line. Extremely accurate matching is not essential.

Similarly, we have to couple the feeder to the aerial. This may be done in various ways. For a simple half-wave aerial we can tap the feeder across a small portion of the aerial itself, as shown in Fig. 18. As already explained, the aerial may be considered as a tuned circuit, and by tapping across an appropriate length we can arrive at a suitable impedance.

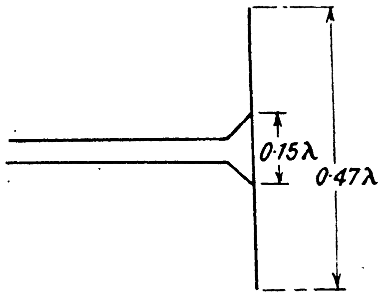

FIG. 18. TERMINATION AT AERIAL FOR UNTUNED FEEDER

To take a practical example, consider a half-wave aerial for a 30-metre transmission. The length would be approximately 15 metres, and if we assume the wire to be 0·25 cm. diameter we can evaluate L and C from the expressions given on page 43. If R is taken as 150 ohms, the effective impedance $4L/CR$ becomes 6 870 ohms.

We require a tapping which will give us about 600 ohms (assuming a parallel wire feeder of usual dimensions). Thus

$$4h^2/\lambda^2 = 600/6\,870,$$

whence

$$h = 0{\cdot}15\lambda \text{ nearly.}$$

Alternatively we can connect the feeder, through a suitable transformer, between one end of the aerial and earth. The base resistance of a half-wave aerial is L/CR (page 44), which in the above instance would be 1 670 ohms nearly. By using a transformer with a step-up of 1·66 to 1, we should match our line satisfactorily.

As previously mentioned, all these calculations are only approximate, since they are based on assumed values of aerial

constants, particularly resistance, and they serve mainly as an indication of the order of transformer or tap required. The final adjustment is made by trial and error.

Matching Lines.

A short length of feeder may be used to match a transmission line to the aerial instead of a conventional transformer. Such feeders are known as matching lines, and are of two types.

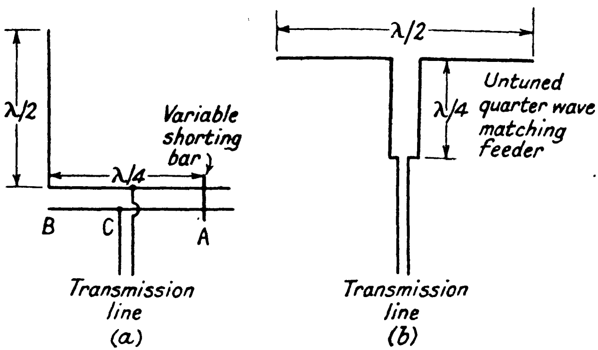

FIG. 19

The impedance of a tuned feeder will obviously vary from point to point. At a voltage node the impedance (neglecting resistance) is clearly zero, while at the current node $\lambda/4$ away the impedance is infinite, since there is no current. A quarter-wave tuned feeder can thus be used as a variable-impedance transformer.

Fig. 19 (*a*) shows such an arrangement. The impedance at A is zero, while at B it is infinite and equal to that of the half-wave aerial, which also presents an infinite impedance at its extremity. The transmission line is tapped across a suitable intermediate point at which the feeder and line impedances are equal. The tapping may be calculated from the expression

$$Z = \sqrt{(L/C)} \cot \omega \sqrt{LCx}$$

where x is the distance from A.

In practice it would be adjusted by trial and error, while the length AB would also be set for optimum results by the variable shorting bar at A, to allow for the fact that the impedance at B is not infinite (because of the resistance in circuit) so that a full quarter-wavelength is not required.

Fig. 19 (*b*) shows an alternative arrangement. Here the same quarter-wavelength of feeder is used, though it is employed as an untuned feeder, and the dimensions are so chosen that the impedance of the matching line $R_m = \sqrt{R_1 R_2}$, where R_1 and R_2 are the impedances of the aerial and the transmission line respectively. This method can only be used successfully if R_1 and R_2 are not widely different.

Practical Forms of Feeder.

The simplest type of feeder consists of two parallel wires, suitably suspended and kept apart. The wires may be run on cross arms fixed to poles, or where the length of feeder is short, it is sufficient merely to insert spacers at suitable intervals and allow the feeder to hang freely between aerial and receiving point.

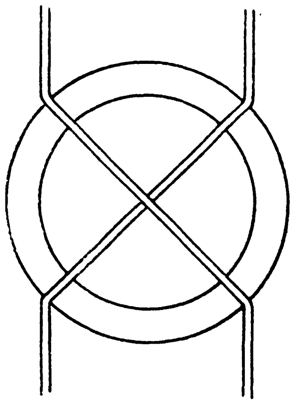

FIG. 20. TRANSPOSITION SPACER

For receiving aerials, transposition is sometimes resorted to, the wires being periodically changed over. This is done to minimize any pick-up of interference from local sources, since any voltage induced in one section is then cancelled by that in the next section, where the pick-up is reversed due to the transposition. Fig. 20 shows a combined spacing and transposition insulator.

It is indeed possible to use simple twisted flex as a feeder, particularly if of the untuned variety, and for television and ultra-short wave reception this is commonly done. The length of feeder in such cases is, of course, only a few wavelengths and the losses are relatively unimportant.

It will be obvious that the spacing between the wires must

have some effect on the efficiency of the feeder, and we shall discuss this point shortly. It is desirable first, however, to consider the question of characteristic impedance, which is equivalent to $\sqrt{(L/C)}$.

The inductance of a pair of parallel wires (assuming the far end closed to complete the circuit) is

$$L = 9{\cdot}21 \times 10^{-3} \log_{10} (d/r) \ \mu\text{H. per cm.}$$

$$\left.\begin{array}{l}\text{where } d \text{ is the spacing of the wires} \\ \text{and} \quad r \text{ is the radius}\end{array}\right\} \text{in similar units}$$

The capacitance between them is

$$C = \frac{10^{-4}}{828 \log_{10} (d/r)} \ \mu\text{F. per cm.}$$

Hence the characteristic impedance, $\sqrt{(L/C)}$, is

$$Z_0 = 276 \log_{10} (d/r) \text{ ohms.}$$

If $d/r = 100$, this is approximately 550 ohms, but the change is small, and a variation of 3 to 1 either way in d/r only changes Z_0 by some 25 per cent.

Concentric Feeders.

An alternative type of feeder which is often used, mainly in the untuned form, is that employing one wire running along the centre of a tube, the tube forming the earth return. Such a feeder has the advantage that radiation losses are small and the mechanical construction is robust, but it is, of course, more expensive than the simple open wire feeder. The inductance is given by

$$L = 4{\cdot}61 \log_{10} (r_2/r_1) \times 10^{-3} \mu\text{H. per cm.,}$$

r_1 and r_2 being the radii of the inner and outer conductors respectively, while the capacitance

$$C = \frac{10^{-5}}{41{\cdot}45 \log_{10} (r_2/r_1)} \ \mu\text{F. per cm.,}$$

so that the characteristic impedance is $138 \log_{10} (r_2/r_1)$.

Losses in Feeders.

Loss inevitably arises in a feeder due to the resistance of the conductors, leakage and dielectric loss at the insulators,

and radiation. The latter two are not readily calculable, but they should in any case be small. Conductor loss is the most important factor, and this is only serious in long feeders.

With an open feeder, the radiation loss decreases as the wires are brought closer together. On the other hand, the characteristic impedance falls, and hence the feeder has to carry more current, so that unless the wire gauge is suitably increased, the conductor loss will rise. Generally speaking, a spacing which makes Z_0 between 400 and 600 ohms is the best.

With a concentric feeder, radiation loss is practically nil, so that conductor loss is almost the only consideration. It can be shown that the optimum ratio of r_2/r_1 is about 3·6, the diameter of tube in practice being three to four inches. This results in a characteristic impedance of about 80 ohms, as already mentioned. The "co-axial" cable used for transmission of television modulation is simply a special construction of concentric feeder having this optimum spacing.

Wave Guides.

Conductor loss increases as $\sqrt{f}$ so that the attenuation produced by a feeder begins to be troublesome at the very high frequencies involved in the micro-wave region. This has led to the development of a new system of transmission in which conductors, as such, are not employed at all. Such devices are called *wave guides* and are discussed in Chapter XII.

The reader desiring further information on short-wave aerials and arrays should refer to—

"Phase Relationships in Beam Systems," Wilmotte and McPetrie. *Journal I.E.E.*, Sept., 1928.

"Action of Reflecting Antennae," Palmer and Honeyball. *Journal I.E.E.*, August, 1929.

"Directional Transmitting Systems," Sterba. *Proc. I.R.E.*, July, 1931.

"Short-wave Directive Antennae," Bruce. *Proc. I.R.E.*, August, 1931.

CHAPTER IV

AERIAL ARRAYS

WE saw in the last chapter that simple aerial systems were limited in their application, and that in order to obtain good radiation it was necessary to use tiered aerials, either singly or in groups. Practically all commercial transmissions utilize a number of such aerials arranged to provide a concentration of the energy in one direction. Such systems are called *aerial arrays*, and there are numerous types in use.

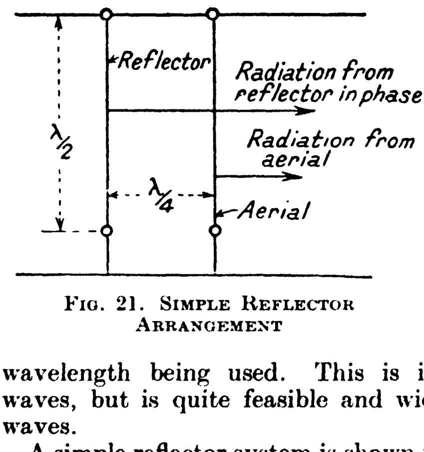

FIG. 21. SIMPLE REFLECTOR ARRANGEMENT

Reflectors.

This concentration of the energy is possible by reason of the fact that the wavelength is short. Any form of reflector must have dimensions of the same order as the wavelength being used. This is impracticable with long waves, but is quite feasible and widely adopted with short waves.

A simple reflector system is shown in Fig. 21, where we have a half-wave aerial with a reflector wire also half a wavelength long, spaced a quarter of a wavelength away. The aerial is supplied with current and induces current in the reflector wire. This current will be lagging by 90° in accordance with the usual a.c. laws, and will in its turn radiate wireless waves.

Consider the wave radiated from the reflector from left to right (i.e. in the direction of the aerial). It starts 90° behind the aerial radiation, but by the time it reaches the aerial a time equivalent to a 90° phase shift has elapsed, so that the

two radiations are in phase. Thus the radiation from the reflector assists that from the aerial.

In the opposite direction, however, the reverse effect will happen, and the two radiations will cancel out. Thus we obtain directional transmission, the radiated field strength being distributed somewhat as indicated in Fig. 22. This is called a *polar diagram*, and represents the field strength in different directions, varying from a maximum in the forward direction, and falling to zero in the backward direction.

It will be noted, however, that there is still considerable radiation at the sides, whereas it would be much more satisfactory if we could concentrate the radiation almost entirely in the direction we require. This can be done by using several aerials side by side, each provided with its own reflector. Let us consider, first, the effect of radiation from two aerials without reflectors.

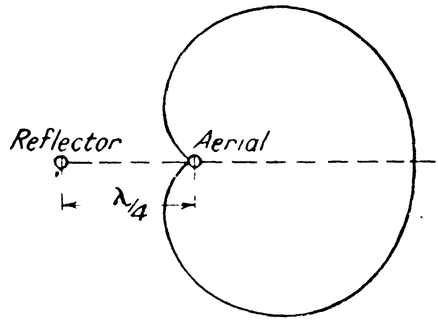

FIG. 22. POLAR DIAGRAM OF RADIATION WITH QUARTER-WAVE REFLECTOR

When the aerials are quite close together they act as one, but as the separation between them increases it is evident that there must be a phase difference between the radiations from the two aerials in certain directions. This phase difference will be most evident along a line joining the two, and we can draw polar diagrams showing the distribution of the field strength at different angles.

Fig. 23 (*a*) shows the diagram for two aerials spaced a quarter of a wavelength apart, from which it will be seen that there is a definite increase in the radiation on a line at right angles to the two aerials, and a reduction in the direction along the line joining the two.

The difference between this polar diagram and that shown in Fig. 22 should be noted quite clearly. The former diagram was for two aerials (or an aerial with reflector), in which the

currents were 90° out of phase. In the present instance the two aerials are both supplied with current *in phase*, which produces an entirely different effect.

Now, as we increase the spacing between the aerials, the concentration of energy becomes more marked up to half a wavelength, when we obtain a diagram of the form shown in Fig. 23 (*b*). Beyond this the effects become complex, producing diagrams having "rabbits' ears" and altering the general

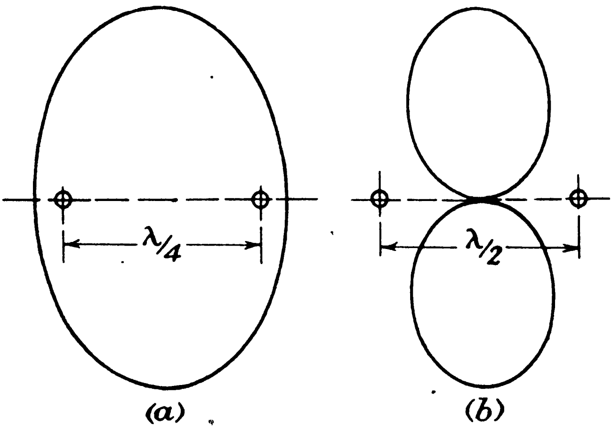

FIG. 23. RADIATION FROM TWO AERIALS IN PHASE

direction of concentration. A typical example is shown in Fig. 24 for two aerials spaced two wavelengths apart. More usually, therefore, the aerials are spaced half a wavelength apart, and if we use several such aerials instead of only two, the concentration becomes more and more intense. Fig. 25 shows a polar diagram for four aerials which will be seen to concentrate the radiation within quite a narrow beam, plus two small subsidiary loops.* The radiation, however, is bi-directional, but we can make it uni-directional by providing

* These polar diagrams can be built up from standard simple diagrams. For example, the four aerial case is equivalent to two groups of two aerials one wavelength apart. By combining the diagram for two simple aerials one wave apart with one for two aerials $\lambda/2$ apart, we obtain the diagram of Fig. 25.

reflectors behind each aerial when the backward radiation is cancelled practically completely.

Still greater concentration of energy can be obtained by using more aerials in line, so that the modern short-wave

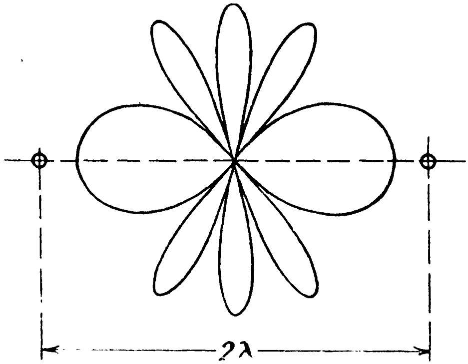

FIG. 24. TOO LARGE SPACING DESTROYS THE CONCENTRATION

transmitting aerial consists of a curtain of wires and reflectors broadside on to the direction of radiation. Fig. 26 shows a Marconi array which comprises a series of tiered aerials sepa-

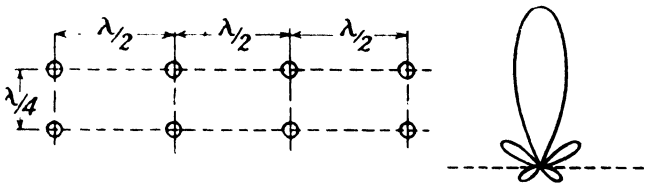

FIG. 25. SIMPLE ARRAY WITH RADIATION DIAGRAM

rated by half a wavelength, while in the rear is a curtain of reflectors at a distance which is sometimes a quarter of a wavelength and sometimes three-quarters. Actually it is found that better results are obtained by having twice as many reflectors as radiators, so that the reflectors are spaced a quarter of a wavelength, and each one comprises two or three parallel wires a few inches apart. This added complexity

is introduced not so much for improving the polar diagram as to give less critical tuning. Fig. 27 is a photograph of an actual installation.

Aperture.

The total width of the curtain at right angles to the direction of radiation is usually called the *aperture*, and may be

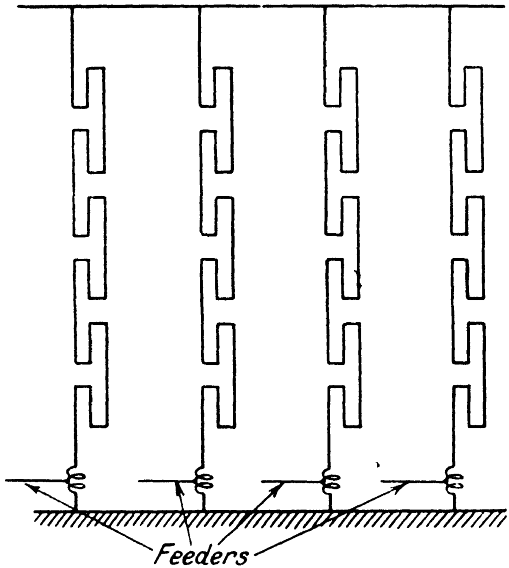

FIG. 26. MARCONI AERIAL ARRAY

anything from two to eight wavelengths. The higher the aperture, the greater the concentration, but, of course, the greater the expense. The aerials are slung from triatics suspended between self-supporting towers, very elaborate cross-bracing being used to obtain rigidity of the structure, and thereby prevent relative movement between the aerials and the reflectors as far as possible. Counterweights are provided at the bottoms of both aerials and reflectors to keep them vertical.

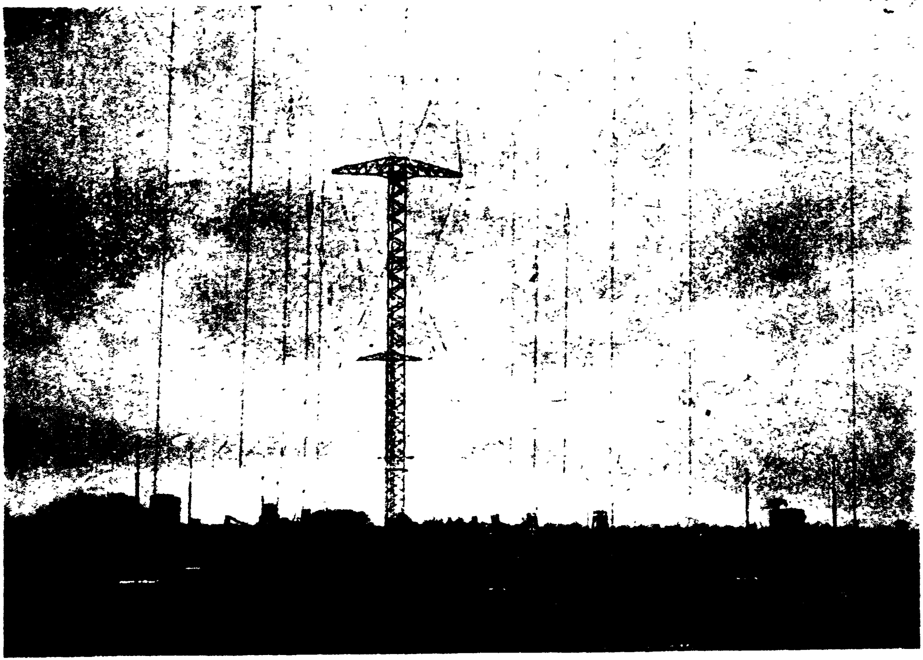

Fig. 27. View of the Aerial System at the Grimsby Short-wave Station
(*Marconi's Wireless Telegraph Co.*)

Feeding the Aerials.

It is necessary that all the aerials shall be supplied with current in the same phase. This is done by means of feeders all radiating from the transmitter. These feeders are of the transmission line type, and the whole system is made as symmetrical as possible, as illustrated diagrammatically in Fig. 28.

There is a main feeder line from the transmitter to a point in the centre of the array. From this point a number of subsidiary feeders radiate, each one being further subdivided, so

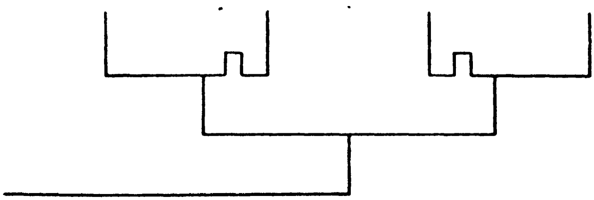

FIG. 28. ARRANGEMENT OF FEEDERS TO SECURE CORRECT PHASE

that each aerial is ultimately supplied with current. The length of each feeder is made the same, being artificially zigzagged if necessary to provide the extra length, so that the voltages at the ends of the feeders are all in phase.

The junctions of the various feeder points have to be made through transformers to maintain the correct impedance, as was explained in the last chapter.

Swinging the Beam.

One advantage of this type of array is that the direction of radiation can be varied within small angles by altering the phase of the current at different ends of the curtain. If, for example, one end is supplied with current before the other, the interference between the radiations will obviously be modified, and actually it will bias the beam to one side. It is essential, of course, that the variation of phase shall be gradual throughout the aerial, and this is accomplished by adjusting the lengths of the feeders so that there is a gradually increasing phase lag in the aerial currents from one side of the array to the other.

The alteration in the length of the feeders is obtained by inserting extra sections of the required lengths, these being bent back on one another so that the two ends come together, and for this reason they are often known as *trombones*.

Standing Wave Arrays.

The array just described uses entirely separate aerials fed in the correct phase, but alternative forms have been devised,

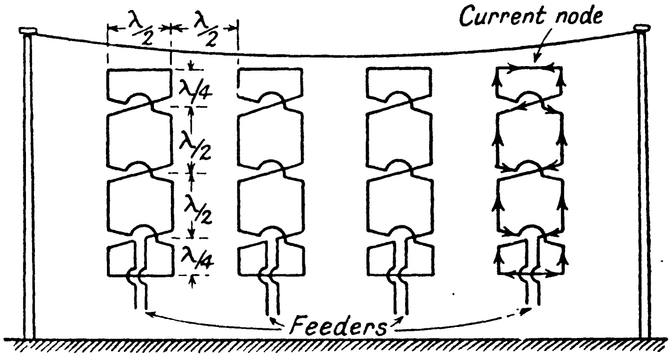

FIG. 29. "STERBA" AERIAL ARRAY FOR SHORT-WAVE TRANSMISSION

using more or less continuous lengths of wire so constructed that standing waves are produced which add up to give the necessary radiation in the forward direction. One of these is the *Sterba array* used by the International Telegraph and Telephone Corporation of America. This array consists of a number of groups arranged as shown in Fig. 29, wherein the successive half wavelengths are alternately vertical and horizontal.

The vertical components thus carry currents in phase, while the horizontal components cancel out. Each unit replaces two individual aerials of the Marconi array just discussed, and the complete array comprises a number of these units spaced half a wavelength apart, each fed with its own feeder. As before, the various units must all be in phase with one another. A

similar construction spaced a quarter of a wavelength to the rear provides the reflector.

The extension of this idea is to be found in the *T.W. array* of the British Post Office, named after the designer, T. Walmsley. In this, the whole aerial is continuous, but is folded back on itself in such a way that the vertical components of the current add up, while the horizontal components cancel. This particular aerial is peculiar in that it

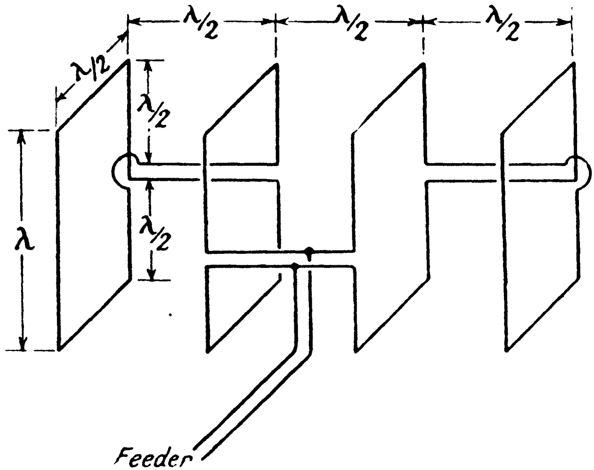

FIG. 30. T.W. ARRAY

has a depth of half a wavelength in the direction of transmission, the currents to the back sections being 180° out of phase with those in the front, and therefore producing additive radiation. The system only requires one feeder introduced at the central point, so that the necessity for accurate phasing of the various sections is obviated.

The radiation is bi-directional, so that the complete aerial still requires a reflector which is provided by a simple curtain of wires a quarter of a wavelength behind the rear section, and hence three-quarters of a wavelength behind the front section.

There are other types of array which need not be discussed

here. In particular, horizontally polarized arrays are sometimes used, consisting of a series of horizontal half-wave aerials so arranged that the vertical radiation cancels out. This gives a horizontally polarized radiation which is just as effective for long distance communication. It is preferably used with a horizontal receiving array, although the rotation and the plane of polarization resulting from the reflection at the Heaviside layer renders this a point of minor importance.

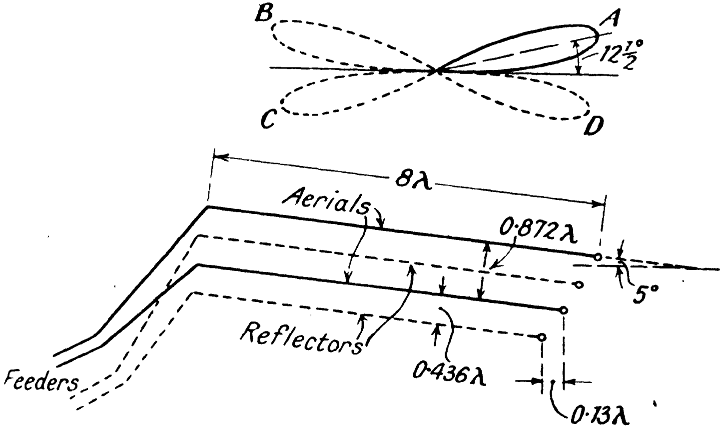

FIG. 31. R.C.A. HARMONIC AERIAL

Long Wire Radiators.

There are other forms of aerial which will concentrate the energy in one direction, and it is sometimes more convenient to use such types. Moreover, a directional receiving aerial is of considerable use, but it is not usually necessary or desirable to employ such a high concentration.

Some of the receiving aerials used in practice are discussed in the next chapter, but there are certain arrangements which are used for transmission which depend upon quite different principles from the broadside reflector types so far considered.

We saw in the previous chapter that a vertical wire several wavelengths long concentrates the radiation in an upward direction and actually, as we increase the length, the effective

radiation makes a decreasing angle with the wire itself, tending in the limit to radiate along the direction of the wire. This effect is used by the Radio Corporation of America in their *Horizontal Harmonic array*, illustrated in Fig. 31.

This consists of a series of wires eight wavelengths long, arranged one above the other. The top wire by itself radiates mainly in a direction 17½° off the axis. This occurs all round the wire, giving a double conical sheath of radiation. The wires are arranged at an angle of 5° with the ground, giving an upward radiation at an angle of 12½°, which is found to be the most favourable for long distance communication.

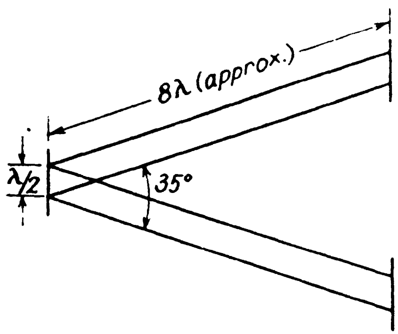

FIG. 32. FOLDED-WIRE ARRAY

A similar wire, a little less than a wavelength below the main wire, is provided with current in opposition to the main current. This wire is actually staggered slightly, and the vertical and horizontal displacements are such that radiation in the directions *B* and *D* are cancelled out. Finally, two reflector wires are provided with current 90° behind the current in the main wires, which cancels out radiation in the direction *C* and increases it in the *A* direction.

It should also be noted that the arrangement of two main radiating wires one above the other, carrying current in opposition, largely cancels out any radiation in directions at right angles to the plane of the array, since in such directions the currents are exactly in phase opposition and the vertical spacing between the two wires does not introduce any phase angle, so that effective cancellation results. The whole aerial therefore directs the more or less concentrated beam in an upward angle of some 12½°, restricted to a few degrees on either side of the vertical plane containing the array.

An alternative arrangement is to use two long harmonic aerials half a wavelength above one another, arranged in the

form of a V, at an angle of 35° and fed in the centre, as in Fig. 32. As before, the radiation is concentrated in directions of $17\frac{1}{2}°$ from the aerials themselves, and it will be seen that this adds up along the line bisecting the V, while the subsidiary radiations are cancelled out by the use of the two wires one above the other, as in the previous instance.

The arrangement is bi-directional, and is therefore provided with a reflector of similar construction situated an odd quarter of a wavelength to the rear. Owing to the length of the aerials this reflector has to be located a considerable distance away, the usual spacing being $2\frac{1}{4}$ wavelengths; and by supplying the reflector with current in quadrature, it is possible to make the aerial radiate in either direction according as the current in the reflector is lagging or leading.

The student who wishes to follow the subject in greater detail should refer to—

Short-Wave Wireless Communication, Ladner and Stoner. (Chapman & Hall.)

"Polar Curves of Extended Aerial Systems," E. Green, *Wireless Engineer*, Oct., 1927.

"Gain of Directive Antennae," Southworth, *Proc. I.R.E.*, Sept., 1930.

"Beam Arrays and Transmission Lines," T. Walmsley, *Journal I.E.E.*, Feb., 1931.

"Development of Aerial Arrays," Carter, Hansell and Lindenblad, *Proc. I.R.E.*, Oct., 1931.

CHAPTER V

RECEIVING AERIALS

THE aerials used for short-wave reception depend upon the service required. Results can be obtained by a simple wire of the type used for ordinary broadcast reception, and indeed for the reception of telephony broadcasts for entertainment purposes this is usually all that is done. A modern home radio receiver almost invariably includes a short-wave range which enables reception to be obtained from all over the world with comparative ease under good conditions.

It is desirable to use a fairly short aerial for reception, since if the length is comparable with half a wavelength, difficulty may be experienced due to the uneven distribution of current of the aerial itself, but if any doubt exists on this score it is a simple matter to insert a shortening condenser, i.e. a simple series condenser of about $100\mu\mu$F. capacitance, which can be adjusted until satisfactory reception is obtained.

If the aerial is too long, the current will distribute itself in a standing wave, the length of which will vary with the wavelength being received, and the effect will be that certain wavelengths are received quite satisfactorily, while others appear "dead." The introduction of a series condenser will alter the current distribution at the receiving end and enable reception to be obtained on wavelengths which were previously poor. It is more than likely, however, that with this condenser in circuit, dead spots will be found on some other part of the range; and, in fact, it is preferable to use a really short aerial not much more than 20 ft. long.

Directional Reception.

For anything but simple home reception, however, it is an advantage to pay more attention to the receiving aerial, and the type of aerial depends upon the conditions. In the case of point-to-point transmission operating on the same wavelength the whole time, it is possible to construct directional

receiving aerials on similar lines to the beam arrays used for transmission. This results in an appreciable increase in energy, for the aerial system is distributed over an appreciable frontage, and the individual wires extract energy from the advancing wireless waves and add their respective contributions in the correct phase to give a greatly augmented signal. Actually the improvement obtained from a beam receiving aerial approaches that of the same aerial used as a transmitter, so that by having beam aerials at both ends a very marked improvement over the simple non-directional type of transmission can be obtained. The papers by Green and Southworth quoted at the end of the last chapter contain some useful information on this point.

A further considerable advantage accruing from the use of beam reception is the reduction of interference. The restricted direction of reception will tend to eliminate unwanted stations and thereby assist the tuning circuits in the receiver. With a commercial receiver, however, operating on a single wave, it is possible to provide highly selective operation, and the question of interference from other stations is not so important as the problem of atmospheric disturbances. These atmospherics, or *X's* as they are sometimes called, are in the form of innumerable sharp impulses of short duration, which provide a background to the reception. They are the result of atmospheric discharges, and are actually very rapidly damped wireless waves. Their severity varies from time to time and season to season. Under good conditions they may be quite negligible in comparison with the wanted signal, while at other times they may produce a continuous succession of crashes which entirely blots out the required station.

Now, since these atmospherics are actually wireless waves themselves, they cannot be eliminated by any tuning device. Their sharp duration produces what is called *shock excitation*, and this forces its way through the receiver irrespective of the tune. Moreover, since the atmospheric disturbances are more or less evenly spread over the whole gamut of wavelengths, we are bound to obtain some interference wherever we happen to set our tune, although there seems to be evidence that atmospheric disturbances are more severe on long and medium wavelengths.

Our only remedy, therefore, is, firstly, to restrict the acceptance band of the receiver as much as possible, so that it only receives the modulation required for the intelligence to be communicated; and, secondly, to use directional methods. The majority of the atmospherics arrive from a considerable distance and are thus not directional in character, so that if we can restrict our reception to a fairly narrow beam of, say, 20°, instead of receiving equally well throughout the whole 360°, we should automatically reduce the atmospheric interference 18 times. While the results in practice do not quite obey this simple mathematical theory, there is a definite improvement from the use of directional reception.

The Beverage Aerial.

Sharp beams, however, can only be used for special services, and it is more usual for a commercial reception to employ semi-directional arrangements, which definitely receive better from the wanted direction than from any other, but are not confined to one particular wavelength for their operation. One of the earliest of these was the *Beverage Wave Aerial*, which is particularly suitable for short waves. It consists of a long wire stretched across the surface of the ground at a height of some 10 to 20 ft. and arranged approximately along the line from which reception is required. A horizontal wire, of course, will extract no energy from a vertically polarized wave, but in practice quite satisfactory reception is obtainable for two reasons.

In the first place, the "feet" of the wireless wave drag slightly, due to the resistance of the Earth, so that the electric field at the surface of the ground is slightly inclined, as shown in Fig. 33. Secondly, the plane of polarization, particularly with short waves, is almost invariably slightly twisted as a result of reflection in the upper atmosphere, and the net result is that there is quite an appreciable component of the wave front which induces voltage in the wire. Now, when the wave strikes the aerial first it induces a voltage which travels down the wire with a velocity practically equal to that of the wireless wave itself. At the same time, the wireless wave is travelling along the wire and inducing further voltages in the succeeding portions. It will be clear that these successive

voltages are reinforced by the travelling wave which has arrived from the beginning of the wire, so that the effect is cumulative and the voltage builds up as the wave travels along the wire until at the far end it is quite considerable. This voltage is then fed to a receiver through a matched transmission line in the ordinary way.

A signal arriving from the opposite direction will obviously build up a similar voltage which will reach its maximum at the point A, but by connecting a resistance to earth at this point and making this equal the surge impedance of the line, any such voltage is completely absorbed without any reflection, so that we have a uni-directional system. It will also be clear that waves arriving from any direction other than that along which the aerial is erected will fail to produce the necessary cumulative effect, and a fairly sharp directivity results.

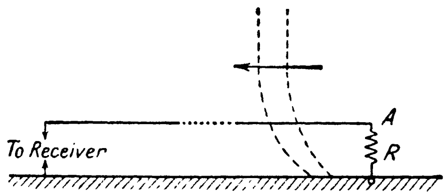

FIG. 33. SHOWING HOW WAVE DRAG IS USED IN THE BEVERAGE AERIAL

The aerial has to be several wavelengths long, which has prevented this form of aerial from coming into considerable use on medium and long waves, but is no serious handicap on short wavelengths.

Zigzag Array.

Another form of wave aerial is the *zigzag array*, which consists of a long wire several wavelengths long bent into a series of zigzags of a quarter of a wavelength side. If such an arrangement is supplied with current at the centre, standing waves will be built up which will have a cumulative radiation from all the vertical sections, and the arrangement, therefore, will act as a form of broadside array of a quarter wavelength height. Since a good radiator is also a good receiver, the converse is quite true, and the arrangement receives from the direction at right angles to its length. A similar zigzag a quarter wavelength behind the first acts as a reflector and

makes the arrangement uni-directional, and since the height is only small, there is little directivity in a vertical direction, which is rather an advantage from the point of view of reception.

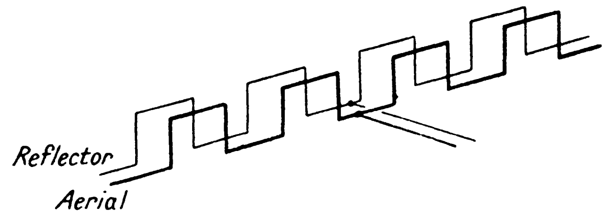

FIG. 34. ZIGZAG ARRAY

Inverted-V Aerial.

Another form of aerial which has the merit of simplicity is the *Inverted-V* type shown in Fig. 35. The sides of the V are one wavelength long and the action is as follows. Consider a vertical wire terminated at the earth end by its characteristic impedance. Voltages will be induced in each part of the wire by an advancing wave, and these voltages in travelling down to the base will all arrive in slightly different phase. It can be shown that if the wire is made one wavelength long, the various component voltages arriving at the base all cancel out, leaving no voltage at all. If, however, we incline the wire into the direction of the wave and make the top half a wavelength ahead, the voltages at the top of the aerial arrive sufficiently in advance of those at the base to compensate for the extra distance they have to travel, and instead of zero voltage we obtain a maximum.

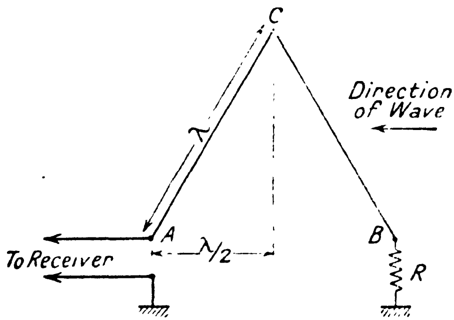

FIG. 35. "V" AERIAL FOR SHORT-WAVE RECEPTION

Instead of terminating in the characteristic impedance, we lead the voltage to the receiver through a correctly matched transmission line, while in practical form the aerial is usually slightly modified by completing the V with a further wire running from the top to the ground a further half wavelength ahead, this wire being terminated in a resistance as shown. This modification improves the cumulative action, while any wave arriving from the opposite direction builds up the voltage which is absorbed by resistance R. A wave from the wrong direction or of the wrong wavelength does not build up, so

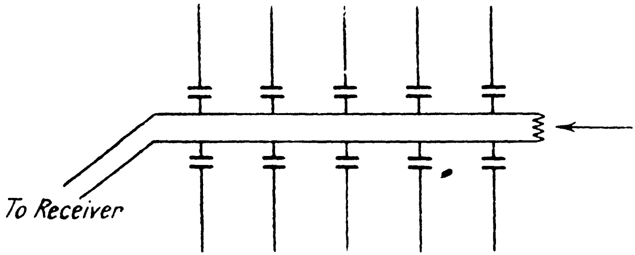

FIG. 36. R.C.A. FISHBONE AERIAL

that a very sharp discrimination both as regards direction and wavelength can be obtained.

Horizontally-polarized Aerials.

The fact that short waves arrive with as much horizonta as vertical component in their wave front has resulted in the use of receiving aerials which only respond to horizontally polarized waves. There does not seem to be much advantage in such systems, except that it is possible to make them several wavelengths wide if desired, which is equivalent to but somewhat simpler than erecting a high aerial for vertically polarized reception.

One of these is the *Fishbone array* developed by the R.C.A. This is shown in Fig. 36. It is rather similar to the Beverage aerial in operation. A series of horizontal half-wave aerials is erected, spaced by a small fraction of a wavelength, usually about $\lambda/6$, and connected to the transmission line. The voltages are induced in each aerial in turn and communicated to

the transmission line. As with the Beverage aerial, we obtain a travelling wave which is reinforced by the voltages from the successive aerials as the wave travels along the aerial, while from other directions the building-up process does not occur. As before, the system is made uni-directional by connecting an absorbing resistance across the front end.

The horizontal diamond is another form of aerial, in which the sides of the diamond are made long compared with the wavelength, so that a building-up process is obtained. The arrangement is shown in Fig. 37, and it will be clear that the

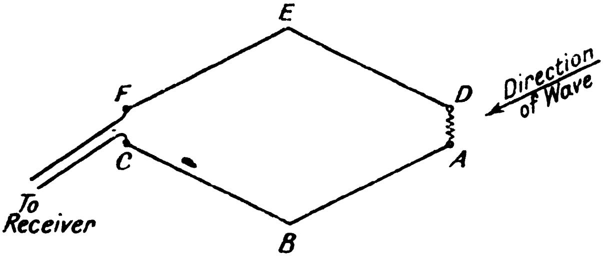

FIG. 37. HORIZONTAL DIAMOND AERIAL

vertically polarized component of the wave has no effect, since it induces equal and opposite voltages in the two sides *ABC*, *DEF*. The horizontally polarized component, however, produces a cumulative effect. Consider a wave arriving from the direction of the arrow. It will induce no voltage in *AB*, which is at right angles to the wave front, but *DE* is at an angle to the wave front, and voltage is built up by a cumulative process as with the Beverage aerial. As the wave sweeps over the second part of the array, it is the side *BC* which is effective, *EF* having no voltage induced, and merely serving to transmit the voltage already produced. Thus at *CF* we have the cumulative effect of the whole aerial, while a resistance across *AD* absorbs any voltage built up by waves in the reverse direction.

The length of the sides is made several wavelengths, while the angles of the sides depend partly upon the length of side and partly on the wavelength. This form of aerial is effective against local interference, which is usually vertically polarized.

Diversity Reception.

The aerials we have discussed so far have been designed to make the greatest possible use of the signal available, but as we have seen, short-wave signals suffer considerably during their reflection from the upper atmosphere. In particular, due to rotation of the plane of polarization, the signal received on any one aerial may completely vanish for short periods lasting from a fraction of a second up to several seconds.

Any normal variation in strength can be overcome by the use of automatic gain control on the receivers, but such systems work on the principle of reducing the amplification when the signal is strong, so that the overall result is no louder than when the signal is weak. No form of automatic control of this type can operate if there is no signal present, so that it is useless against a complete fade-out.

It is found, however, that when an aerial in one position is suffering from a fade, another aerial a few wavelengths away may be receiving quite satisfactorily, an effect which is quite understandable, since a difference of a wavelength or so in the distance travelled by the wave may enable the plane of polarization to rotate from horizontal to vertical.

This has led to the introduction of what is called *diversity reception*, in which several aerials are erected a few wavelengths apart. They are all connected by correctly terminated transmission lines to the receiving room, where the signals are mixed. It is, however, impracticable to mix them straight away, for the phase of the modulation on the different aerials is not necessarily the same, and therefore each aerial is fed into its own radio-frequency amplifier, up to and including the detector stage, and then the outputs from all the detectors are mixed and passed through a common audio-frequency amplifier.

By making the detectors obey a square law characteristic, we automatically ensure that the greater part of the output comes from the receiver which is handling the greatest signal at that moment, and this automatic shifting of the load from one aerial to another operates quite well in practice. The inevitable distortion which accompanies fading, often due to selective fading, is partly reduced by diversity reception, though often not entirely overcome.

For telegraphic communication, use is sometimes made of what is called *frequency diversity*. The signal is modulated with an audio-frequency tone and several receivers are used tuned to slightly different frequencies. Thus some part of the total modulation band is received on one or other of the receivers, and sufficient signal is obtained to enable communication to be carried on except under really severe conditions.

Musa System.

Messrs. Friis and Feldman of the Bell Telephone Laboratories have developed a system known as a Multiple Unit Steerable Antenna, or "Musa." Detailed investigations have confirmed that signals arrive at the receiving point by multiple reflection from the upper atmosphere as explained on page 18. Due to circular polarization, the arrival angle of the wave having the best field strength is continually changing, and it has therefore been necessary to make the vertical directivity of the receiving aerial system sufficiently broad to cover a range of angles.

The Musa arrangement employs a series of fixed rhombic (horizontal diamond) aerials arranged in line and each fed to a phase shifter. The outputs of the phase shifters are combined and fed into a receiver. By operation of the phase shifters, which are all ganged together, it is possible to vary the vertical directivity of this array. Moreover, the beam is very much sharper than with a simple arrangement, so that if the angle can be correctly chosen to coincide with the incoming signal, a marked improvement in signal-to-noise ratio will result.

Actually, the output from the six aerials is fed into three separate phase shifters operating at parallel. The first two select two vertical angles separated by a few degrees, and the outputs from these two channels are subsequently combined. This, therefore, provides a diversity reception, an audio-frequency delay being introduced into the channel receiving at the lower angle, the value of this delay being adjusted until the two audio-signals are in phase. The audio-frequency phase is checked by means of a cathode-ray monitor tube, which produces the customary phase ellipse. Correct adjustment causes this ellipse to resolve itself into a single inclined line,

while if the delay is incorrect, a maze of irregular circles and ellipses is seen.

It remains to ascertain which is the correct angle at any given time, and this is done by the third channel on which the phase shifter is continually rotated by a motor. The horizontal deflecting plates of a second cathode-ray tube are also linked with this motor drive in such a way that the spot moves horizontally across the screen once for each complete revolution. The vertical plates are operated from the output of a receiver so that at some part of the travel across the screen the spot will deflect sharply in a vertical direction, and the position of this peak will indicate the optimum angle of signal arrival at that time. The phase shifters of the other two units are therefore set at this angle. As conditions change, the position of the peak on the monitor tube varies, and the phase shifters are altered in conformity. Since the change of angle is relatively slow, this is a comparatively simple matter.

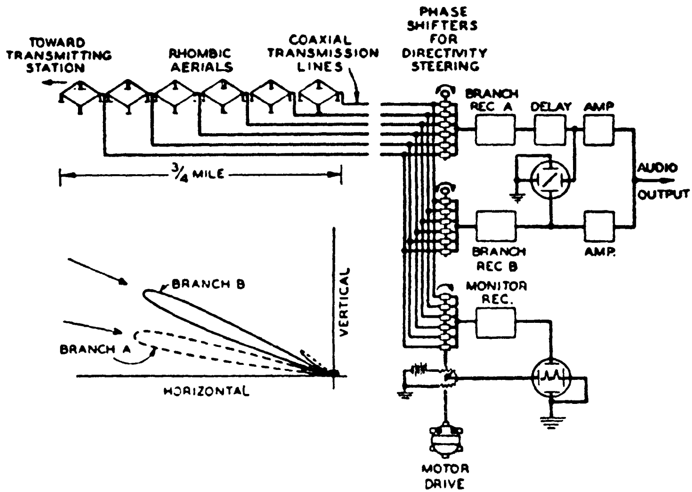

FIG. 38

A skeleton diagram of the arrangement is shown in Fig. 38, while the reader who requires further particulars should refer

to a paper by Ralph Bown before the Wireless Section of the I.E.E. (*Journal I.E.E.*, Vol. 83, p. 395).

It is claimed that in the experimental system at Holmdel, New Jersey, an improvement in signal-to-noise ratio of 7 to 8 db. is realized, and that the reduction in distortion is frequently quite marked.

Polarization Diversity.

Use is occasionally made of the fact that, when a signal has faded out on a vertically polarized aerial, it can usually be quite strongly received on a horizontally polarized one, and

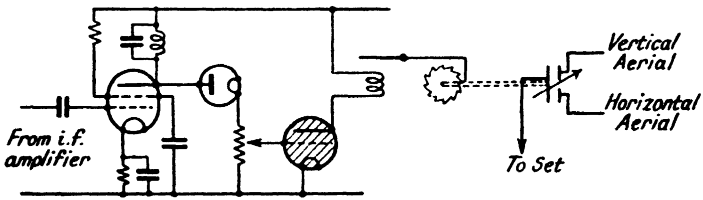

FIG. 39

vice versa. The normal principles of diversity reception can be applied to this arrangement equally well.

McMurdo Silver has described an automatic arrangement suitable for simple receivers which employs this principle.* Two aerials, one vertically and the other horizontally polarized, are coupled to the receiver through a differential condenser, the spindle of which is driven by a relay and ratchet mechanism giving eight distinct positions. In this way, by rotation of the shaft, signals can be picked up to varying extents from either or both aerials. Owing to the fact that the plane of polarization is continually rotating, it is therefore possible to follow the changes by a corresponding rotation of the differential condenser.

Signals are fed from the i.f. amplifier of the receiver on to an additional i.f. stage feeding a rectifier, and the carrier voltage developed is used to bias a gas relay and to hold it non-conducting (Fig. 39). If the signal drops below a certain critical value,

* *Wireless World,* 2nd March, 1939.

however, the bias on this relay decreases, causing the valve to fire and operate the mechanical relay in the anode circuit, which causes the aerial (differential) coupling condenser to rotate one notch. This will restore the signal strength to its former value, and reception continues until the signal strength again falls below the critical value.

CHAPTER VI

SHORT-WAVE TRANSMITTERS

THE production of the high-frequency oscillations fed to the aerial system can now be considered. These oscillations are generated by suitable valve oscillator circuits, and, if necessary, are amplified by further valves before being supplied to the aerial.

The simplest form of circuit is that shown in Fig. 40, which illustrates a self-oscillating system containing a tuned circuit in the anode and a coupling coil in the grid circuit connected in such a direction as to maintain oscillation. A parallel-fed anode circuit is usually employed mainly to avoid any h.t. voltages on the tuned circuit.

Now the oscillations developed in this circuit may be transferred directly to the aerial circuit, but this is only done in the case of low-power transmitters where simplicity is the essential consideration. In the first place, it is by no means easy to generate a large amount of oscillating power in one stage. Secondly, constancy of frequency is most important on short waves, and this again is difficult to obtain with large powers. For both these reasons it is more usual to generate relatively small power oscillations and to amplify these subsequently.

There is a further difficulty in connection with the direct-coupled oscillator, namely, that the power which can be transferred from the oscillating circuit to the aerial is limited. Since both the valve anode circuit and the aerial are tuned, they constitute a coupled-circuit system which has more than one mode of oscillation, and if the coupling is increased by more than a certain amount, the operation becomes unstable and the oscillation may change from one mode to the other indiscriminately. Since the two oscillations are of different frequency, this is clearly an undesirable state of affairs and is to be avoided.

The critical coupling at which this instability occurs is given by the expression

$$M = (R_2/\omega L_2)\sqrt{(L_1 L_2)},$$

where L_1 and L_2 are the inductances of primary and aerial circuits respectively and R_2 is the total secondary resistance (including radiation resistance). This mutual inductance is

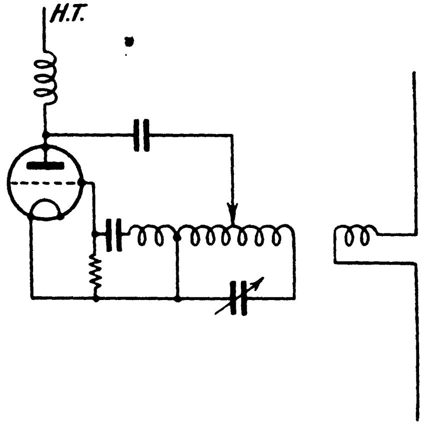

FIG. 40. SIMPLE DIRECT-COUPLED TRANSMITTER

appreciably less than that for optimum energy transfer, which means that the full power cannot be extracted from the circuit. For the derivation of this expression for critical coupling, the reader should refer to *Modern Radio Communication*, Vol. II.

Still another objection to the simple direct-coupled transmitter is that the aerial constants have an appreciable effect on the frequency, so that any movement of the aerial, due to wind or other causes, produces a frequency variation which is to be avoided at all costs.

The usual procedure, therefore, is to generate oscillations with a small low-power oscillator and to follow this with a chain of amplifier valves. We shall discuss the technique of this process in a moment, but first of all we will consider the problem of obtaining the greatest efficiency in the oscillator.

High Efficiency Working.

The ordinary oscillator or amplifier operates under what are known as "Class A" conditions. The valve is biased to the mid-point of its working characteristic, and the variations of grid voltage cause the anode current to swing about this point, between the limits of zero current and twice the normal, though actually the variation is only some 80 per cent of this theoretical maximum, due to the curvature of the characteristic.

The principal feature of this condition is that the mean anode current remains unchanged. Hence there is a continuous drain on the h.t. supply amounting to the product of the h.t. voltage and the mean anode current, and we have to convert as much as possible of this power into actual oscillating power. Now it is easy to see that the maximum possible efficiency with such an arrangement is 50 per cent. The maximum voltage swing is from the normal h.t. value to zero, and similarly for the current swing. These are peak swings, and the r.m.s. value is obtained by dividing by $\sqrt{2}$ in each case. Hence the oscillating component of the anode power is $(v_a/\sqrt{2})(i_a/\sqrt{2}) = \frac{1}{2}v_a i_a$, which is one-half the steady power.

The voltage and current in the oscillating circuit may, of course, be much greater than this, but since the oscillating voltage and current are in quadrature they do not represent the consumption of power, but merely an alternate storage of energy in magnetic or electrostatic form. The power in the circuit is represented by the relatively small component of the current which is in phase with the voltage, and this power loss has to be made good by the valve if the oscillation is to be maintained, so that the oscillating component of the anode power is equivalent to the power dissipated in the oscillating circuit.

Impulsing.

Better efficiency can be obtained by arranging that the valve only draws anode current at certain parts of the cycle. The peak current which does flow will obviously have to be somewhat larger, but due to the idle periods, the mean current is appreciably less. There are two ways of obtaining this

effect. One is to bias the valve nearly to the cut-off point, so that on the positive half-cycle anode current can flow, but no current can occur over the negative half-cycle.

This method of operation is known as "Class B" working, and has the advantage that the amplitude of the oscillation produced is still reasonably proportional to the grid swing, for if we make our grid excursion more and more positive, the anode current impulse gets greater and greater, more power is supplied to the oscillating circuit, and hence a greater oscillating current is built up. This point is of importance in driven circuits (i.e. circuits using amplifiers following the oscillator), as we shall see later.

The third condition is "Class C" working, in which the valve is biased *beyond* the cut-off point so that the anode current only flows at the positive peaks of the grid swing. At such points there is a very large pulse of anode current which is sufficient to maintain the oscillation, but for the rest of the cycle the anode current is zero. With such a system a practical efficiency of between 80 and 90 per cent can be obtained, as against about 40 per cent for a Class A circuit. Class B gives an efficiency in between these two.

Transmitting Valves.

It will be clear that the operation of an oscillating valve is quite different from that of an amplifier. The valve must be capable of giving large peak anode currents anything from six to ten times the average value, while, in addition, the valves are almost invariably operated so that, at the peak, the grid is appreciably positive.

Valves intended for oscillators are therefore specially designed to withstand these conditions, but there are certain limits which must be observed. Fig. 41 shows the characteristics of the simple triode oscillator, and it will be seen that there are two limits marked thereon. The first of these is the *emission limit*, which is the maximum current which the valve will give, even momentarily, without being damaged; the second is the *limiting edge*. It will be seen that as we make the grid less negative, the characteristics shift to the left. This continues as the grid is made positive up to a certain point, but the characteristics here begin to crowd up and a

limit is reached, when further positive bias makes little difference to the anode current.

We have to work our valve up to these limits in order to obtain the maximum energy from it, and this is a matter of choosing the correct anode impedance. In the example shown, for instance, we have assumed an h.t. voltage of 1 000, and at this point we assume that the valve is passing no anode current. The maximum efficiency is obtained if, during the operation, the anode current runs from this point up to the intersection of the limiting edge and the emission limit, since this will give us the maximum possible change of both anode current and anode voltage. In other words, our load line must be as shown on Fig. 41, corresponding, in the example shown, to a value of about 2 000 ohms.

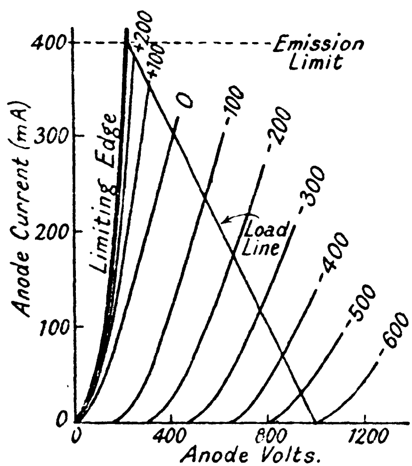

FIG. 41. CHARACTERISTICS OF TYPICAL LOW-POWER TRANSMITTING VALVE

This value may be greater or less than the internal resistance of the valve. With Class A working, the maximum efficiency is usually obtained with the external and internal impedances approximately equal, but this condition no longer applies with impulse working.

Grid Bias.

The necessary bias on the grid is usually obtained by including a condenser and leak in the grid circuit. The grid current which flows when the grid runs positive, charges this condenser negatively, and it will automatically acquire a voltage such that at each grid swing the valve is just able to maintain the oscillation.

If the voltage on the condenser is too great, the oscillation amplitude momentarily drops, so that the next peak grid swing does not run so far into grid current. As the condenser is continuously discharging through the leak, the voltage will fall and correct conditions will automatically be restored.

The choice of condenser and leak values must be such as to allow this automatic adjustment to occur, and this depends on the decrement of the tuned circuit, as explained on page 97. The circuit of Fig. 40 contains a condenser-leak arrangement of the type just described.

It should be noted that a self-biased circuit such as this automatically operates under Class C conditions. For Class B working, a definite and independent bias must be provided.

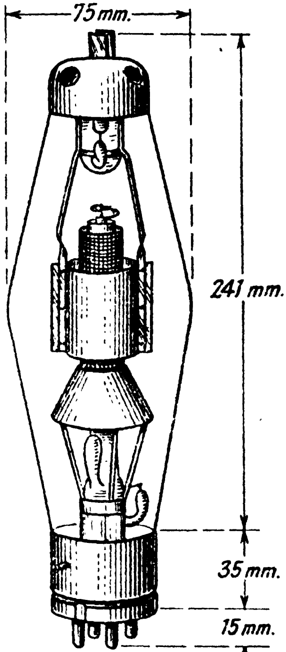

FIG. 42. MULLARD S.G.1 TRANSMITTING VALVE FOR SHORT-WAVE "CLASS B" CIRCUITS

Matching the Valve.

The correct anode load is obtained by coupling the oscillating circuit to the valve with a suitable transformer. The effective dynamic resistance of a parallel-tuned circuit is L/CR, which is usually considerably greater than is required, even with the small inductances used with short-wave working. Consequently a step-down transformer is usually required, and this may conveniently be obtained by tapping the anode of the valve part way down the coil. A tapped coil such as this is, of course, a particular type of high-frequency transformer, and the effective primary impedance can only be calculated accurately from a knowledge of the mutual inductance between primary and secondary. The calculation is easily made by ordinary transformer theory, as explained in *Modern Radio Communication*, Vol. II, but in

practice, since the circuit resistance is usually only approximately known, it is sufficient to obtain a reasonable estimate of the required tapping point, and then to adjust by trial and error until the best results are obtained.

For this purpose it is sufficient to assume the step-down ratio as equal to the ratio of the total turns to the tapped portion, when the effective primary impedance becomes $(L/CR)\,(n_1/n)^2$, where n is the total turns and n_1 the tap.

If a simple oscillator is being employed, we now have to couple it to the aerial. If a direct coupling is being used, the feeder will be coupled to the coil, as already explained in Chapter III, and the mutual inductance will be adjusted not to exceed the critical limit. It should be noted that in such a case the resistance in the matching calculations must include the aerial resistance, which will considerably modify the results.

Drive Circuits.

More usually, however, the voltage developed by the primary oscillator is used to drive a second valve. For powers up to about one kilowatt, this valve may be the final power stage itself. It would be biased approximately to cut-off point by means of an independent bias supply, and the alternating voltage developed by the oscillator would be used to drive the grid up to zero and beyond, the amount of the positive drive depending upon the permissible conditions for the valve and the output requirements generally.

Under such conditions the oscillator is required to develop as much voltage as possible, but this does not mean that this has not to deliver any power. In the first place, the grid will absorb considerable power at the peaks of the oscillation, due to the flow of grid current, which can easily swing up to values of several hundred milliamperes with a one-kilowatt valve, and proportionally more with larger powers. Secondly, the input impedance of the output valve is not infinite, even under conditions where no positive drive is being obtained. The reader will be familiar with the phenomenon known as *Miller effect*, so named after the engineer who first made a detailed study of its action.*

* "The Input Conductance of a Vacuum Tube," by John M. Miller. *Bureau of Standards Bulletin*, No. 351.

When a valve is operating, the voltage developed in the anode circuit due to the amplification of the valve not only passes current through the anode circuit but also sends a current through the internal capacitance of the valve between anode and grid, whence it will flow back to the cathode through the external grid circuit. Consequently, the current flowing in the grid circuit is not only that due to the ordinary static grid impedance (which is usually negligibly small)but is augmented by this feed-back current, which can be considerably greater. The effect, indeed, is as if we had connected an additional impedance across the grid and the cathode of the valve, and the nature of this impedance will depend upon the phase of the feed-back current, which in turn depends upon the nature of the load in the anode circuit.

If the anode load is a pure resistance, the reflected grid impedance is capacitive. If the anode circuit is capacitive, the reflected impedance is a resistance; while if the anode circuit is inductive, the feed-back is a negative resistance, i.e. energy will be fed back in such a direction as to increase the grid current (and consequently the grid voltage), and if this condition is allowed to persist it may cause continuous self-oscillation around the valve in question.

This is the reason why triode valves have been superseded by screen-grid types in the r.f. stages of modern receivers, and it is probably the particular form of Miller effect with which the reader is most familiar; but the other possibilities should be borne in mind, as they have an important effect in the operation of drive circuits.

If we assume, for example, that the anode circuit of the valve is correctly tuned so that the anode circuit of the valve behaves as a pure resistance, we then find that the input impedance looks like capacitance. The drive is usually fed on to the grid from the oscillator through a condenser (see Fig. 44), so that the voltage actually developed at the valve will depend upon the relative reactances of the grid capacitance and the coupling capacitance, and if the latter is small a potentiometer action will be obtained, resulting in considerable reduction in the actual voltage applied to the grid of the output valve.

This, of course, may be overcome by making the coupling condenser large; but this only partially overcomes the

trouble, because the grid capacitance of the output valve is effectively shunted across the oscillating circuit, thereby altering the tune, and if a variable range of frequency is required it may well be found that this is quite seriously restricted by the grid capacitance, which has the effect of causing a permanent and fairly large increase in the stray capacitance of the circuit.

If the capacitance in the anode circuit of the output valve is a little higher than normal, as it may well be during preliminary operations, we shall have the effect of a resistance across the

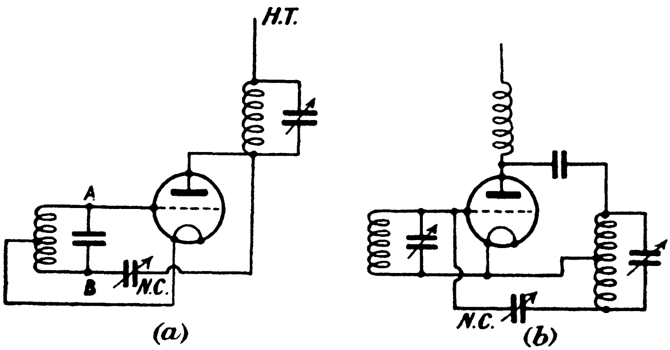

FIG. 43. TWO SIMPLE FORMS OF NEUTRALIZED CIRCUIT

grid and cathode which will absorb considerable power, and it may be impossible to drive the valve to anything like the required extent; while if the anode circuit is inductive (i.e. the tuning capacity too small), power will actually be fed back in the reverse direction and self-oscillation will probably result, so that the output ceases to respond to variations in the drive, either in frequency or in amplitude.

Neutralizing.

There are two remedies available. One is to provide an alternative path from the anode to some point on the circuit which is opposite in phase to the grid, as in Fig. 43 (*a*). Here there are two feed-back paths from the anode, one through the valve and the other through the neutralizing condenser *N.C.* to *B*. But since the grid circuit is symmetrical, the voltages at *A* and *B* are equal and opposite, so that any voltage

introduced at B through $N.C.$ will have the opposite effect to that fed back through the valve. By adjusting the value of $N.C.$, therefore, the valve feed-back may be cancelled out.

Fig. 43 (*b*) shows a similar arrangement, in which the anode circuit is centre-tapped and the neutralizing voltage is not fed from the anode but from a point at which the voltage is opposite in phase to the anode voltage at any instant. This voltage passes a current back to the grid through a neutralizing condenser in opposition to that actually fed back through the valve itself.

Both these forms of circuits are tolerably satisfactory over a limited range of frequency. At short waves, however, the inductance of the leads and the stray capacitances in the circuit quickly become of sufficient magnitude to disturb the true balance, for it is obviously essential that the current fed back through the neutralizing circuit shall be exactly equal and opposite to the valve feed-back, not only in magnitude but in phase. Consequently, as we shall see later, it is necessary to adopt special forms of circuit in which the closest attention is paid to absolute symmetry of the actual leads and disposition of the components in the circuit itself.

Most transmitters are required to operate over a range of frequency, even though this may only be fairly small, and a circuit which may be satisfactory at a fixed frequency will be found impracticable if the conditions are variable.

Screen-grid Valves.

The alternative arrangement is to use screen-grid valves similar to those employed in receiving technique. The reader will be familiar with the general construction of this type of valve, the essential point being the incorporation of a close-mesh grid between the control grid and the anode. This grid is maintained at a positive potential less than that of the anode, and acts as an earth shield which to a large extent prevents the transfer of energy back from anode to grid.

The construction of such a valve for transmitting purposes involves a certain sacrifice of the screening effect, because with a very close screen it is impossible to pass a large anode current through the valve without excessive h.t. voltage. Moreover, since we are handling considerable power in the anode circuit

the dissipation at the screen becomes quite appreciable, and adequate construction must be adopted to permit suitable dissipation of heat.

The feed-back, therefore, is not entirely removed, but it is very largely reduced, and this type of valve is gradually becoming used in commercial equipment, particularly in transmitters requiring anything other than a severely restricted wave range.

Fig. 44 shows the circuit of a 1-kilowatt transmitter using

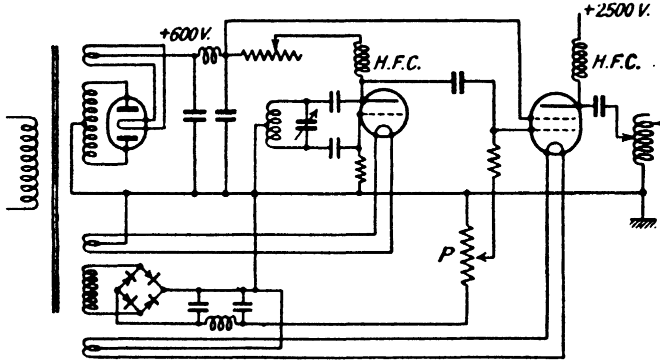

FIG. 44. SKELETON CIRCUIT OF 1-KW. TRANSMITTER

such an arrangement. The drive oscillator is straightforward in character, the drive power being controlled by altering the h.t. to the drive valve. The oscillating voltage is passed on to the output valve, which is provided with a separate pre-set grid bias supply. The potentiometer *P* has to be of low resistance (e.g. 10 000 ohms), so that the grid current drawn on the peaks of the swing will not cause a serious fall in grid-bias voltage.

By adjustment of the grid bias and the drive voltage it is possible to operate this transmitter either as a Class B or Class C circuit, according to requirements. No modulation circuit is shown. This would be added in accordance with the principles discussed in the next chapter.

Efficiency of Driven Valve.

The efficiency of operation of a driven valve is arrived at from similar considerations to those in an oscillator. A Class A

valve is, of course, not used, owing to its poor efficiency, and the valve is usually run in a Class C condition, being biased to cut-off or a little beyond. The oscillating voltage applied to the grid drives the valve up into the positive grid-current region, and, as with an oscillator, a load line can be drawn to determine the conditions of optimum working. The drive is carried up to the emission limit and to the limiting edge, as

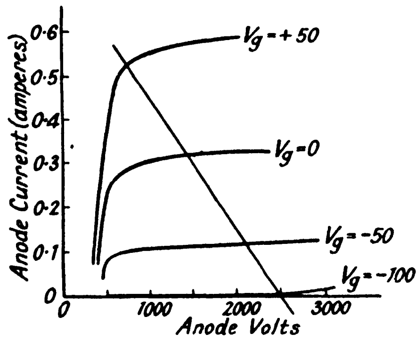

FIG. 45. CHARACTERISTIC OF SCREEN-GRID TRANSMITTING VALVE

before, and with a triode valve the characteristic would be exactly similar to that shown in Fig. 41.

Similar considerations apply with a screen-grid valve except for the fact that the characteristic itself is of different form. It is, in general, not possible to drive a tetrode as far as a triode, for reasons which will be appreciated by reference to Fig. 45, which shows a typical series of characteristics. In Fig. 41 the anode swing is from 1 000 down to approximately 200 volts, i.e. 80 per cent of the h.t. supply, whereas in the case of Fig. 45 the swing is from 2 500 to 750, which is only 70 per cent. The difference arises from the fact that when the anode voltage falls to a value comparable with the screen voltage, the anode current falls off very rapidly and the screen current begins to increase, so that the start of the characteristic is effectively offset by an appreciable amount.

Also, for telephony transmitters it is important that equal increments of grid voltage should cause equal change of anode current. With a tetrode this usually necessitates using a higher

load than the optimum as determined from output considerations only, with consequent loss of efficiency.

Intermediate Amplifiers.

In a high-power transmitter the voltage developed by the oscillator is insufficient to drive the final output stage direct, and it is necessary, therefore, to incorporate intermediate amplifiers. The amplifier valves, however, are arranged on similar principles to those already outlined, each stage having to handle power in order to drive the succeeding valve. A Standard Telephones type 4030 valve, for example, which is a double-ended water-cooled triode, will give an output of 48 kilowatts for telegraphy or 20 kilowatts for telephony at 22 megacycles. Under these conditions the valve requires 1½ kilowatts' drive power on the grid, so that it would have to be preceded by a stage comprising at least two valves similar to those we have just been considering. These drive valves again would require to be supplied with power which might be obtained direct from an oscillator, though it is more likely that it would be obtained from a small power valve which in its turn was driven by a relatively low-power frequency-controlled oscillator using a crystal or some similar means of control.

Bridge Circuits.

As already indicated, triode valves in correctly neutralized circuits are still more common than screened valves in transmitting technique, mainly because they are cheaper in first cost, slightly more efficient, and quite satisfactory under the restricted waveband conditions which obtain in most transmitting practice.

The simple neutralizing circuits as shown in Fig. 43, however, are not satisfactory, owing principally to their lack of symmetry with regard to earth. It has already been explained that complete symmetry is the only way to ensure that the neutralizing shall be correct, not only as regards magnitude but phase.

Considerable use, therefore, is made of push-pull and similar balanced circuits in which absolute symmetry can be obtained. Fig. 46 illustrates a typical circuit devised by C. S. Franklin of the Marconi Company which provides an absolute balance both as regards voltage and phase, even down to very short

wavelengths. At wavelengths below 20 metres it is possible that the reactance of the leads is sufficient to throw the balance out, unless all the leads have been made of the same length, which is not always practicable. The balance, however, may be restored in such cases by introducing in the offending leads —usually the valve leads—a series condenser of reactance

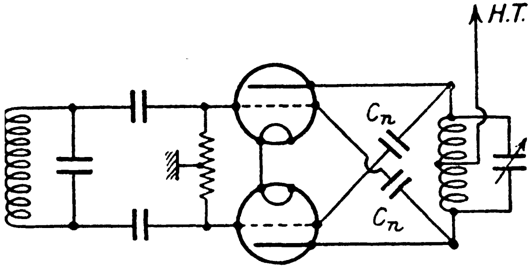

FIG. 46. FRANKLIN BRIDGE CIRCUIT

equivalent to the inductive reactance, which is thereby cancelled out. If the lead is one which has to carry direct current, the condenser will be shunted by a radio-frequency choke which has a much larger reactance than the condenser, and therefore does not upset the operation, but provides the necessary d.c. path.

Valve Losses.

An important consideration in short-wave work is the valve loss. Owing to the small inductances used, it is necessary to employ very small tuning capacitances if any reasonable ratio of L to C is to be preserved. This in turn means that the stray circuit capacitances constitute a considerable proportion of the total capacitance, instead of being only a minor fraction, as is the case with longer wavelengths.

This has two unfortunate effects. The stray capacitances are made up of the valve capacitance and the self-capacitances across the coils, and similar portions of the circuit at a high-frequency potential relative to one another. None of these is a good condenser, so that appreciable losses are set up.

It is important, therefore, to design the circuit so that the dielectric loss in the insulators is as low as possible, so that

the inevitable self-capacitance shall be of reasonably low-power factor and will therefore absorb a minimum of energy. Similar precautions must be taken with the valve. It is impossible to avoid the capacitance effect, and therefore the valve must be designed to carry quite an appreciable proportion of the oscillating current. Obviously, if half the capacitance is in the valve, then half the oscillating current will flow in this portion of the circuit. Consequently, the lead-in wires making the connections through the glass bulb must be thick enough to carry this oscillating current without overheating. The glass itself is a good dielectric and does not occasion much loss, provided that the leads are of adequate size. Trouble occasionally arises due to hot spots on the glass due to eddy currents set up in the metallic deposits formed on the glass during evacuation. This has largely been overcome by using different methods of cleaning up the gas inside the bulb, which do not involve the metallic deposits commonly used in receiving valve technique. If any tendency to hot spots is observed, it may usually be overcome by enclosing the valve in a gauze sheet which produces uniform eddy currents over the whole surface.

Short-wave Circuit Design.

The importance of minimizing dielectric loss has already been mentioned. The circuits used in short-wave transmitters are, therefore, made of skeleton construction wherever possible. The coils, of course, are small and are usually made of copper tube, since the current only flows in the surface of the wire due to the high frequency. Alternatively, a strip conductor may be employed wound with the flat surface parallel to the axis of the coil, thus approximating to a current sheet.

It is particularly important to avoid closed loops in the wiring, since these can absorb energy from the main oscillating circuit and result in quite heavy loss. Wave-changing cannot be carried out by short-circuiting the end turns as is common at longer wavelengths, and it is necessary to change over completely from one coil to another or employ a series-parallel arrangement.

Insulators should only be located at vital points, and the

amount of material used should be the minimum consistent with mechanical strength. The coils are made self-supporting as far as possible, so that occasional insulators only are required. Sulphur-filled fittings are to be avoided, as the sulphur heats up under the influence of the high-frequency dielectric current. Since ebonite contains an appreciable proportion of sulphur, it is not used, the customary materials being of a ceramic nature, such as porcelain, or special brands of steatite known as *frequentite*, *calan*, etc., which have a low dielectric loss at very high frequencies.

Parasitic Oscillation.

A difficulty which arises on short waves is that the transmitter will oscillate on more than one mode, producing what is called a *parasitic oscillation*. The relative strength of this oscillation depends on the conditions, and in severe cases it may swamp the true oscillation altogether. Such oscillations are evidenced by abnormal currents in the circuit, such as high anode feed current, excessive grid current causing the grid to become red hot, etc. Sometimes the circuit shows all the symptoms of a strong oscillation, but the particular frequency cannot be found in the normal range by wavemeter or other measuring device.

Parasitic oscillations usually appear in multi-valve arrangements, though they may be present with quite simple circuits, especially when using valves of high mutual conductance. They are of two general types.

(*a*) *Oscillations near the Fundamental Frequency*. These are usually due to bad lay-out. The leads between the valve and the tuned circuit may form an oscillating circuit of lower resistance than the correct one, and the valve will always choose to maintain oscillation in the easiest mode. A lead too near the framework may form an oscillating circuit due to the capacitance to earth, and there are various other possibilities which would suggest themselves. The remedy is to check the lay-out very carefully for any such subsidiary loops.

(*b*) *Very High-frequency Oscillations*. These are due to self-oscillation in the leads to the valve, and may occur at any part of the transmitter, a particularly noxious example being

in the modulating stages where considerable power may be wasted by parasitics.

On longer wavelengths it is possible to check such high-frequency parasitic oscillations with stopper resistances in the valve leads, but this is not practicable at short waves, and the only remedy is to pay special attention to lay-out, keeping the leads short and avoiding leads of similar length in adjacent stages, so that the natural frequencies of the circuit formed by the leads may be different throughout the chain.

Modulation.

The continuous waves generated by the oscillator must, of course, be modulated before it can communicate intelligence. We may either do this by breaking it up into the dots and dashes of the Morse or some similar telegraphic code, or we may modulate with speech for the transmission of telephony.

For telegraphy transmitters, Class C operation is used, because we only require to start and stop the oscillation with the key at periodic intervals, and when the key is depressed, the amplitude of the oscillation is required to run up to its full value immediately.

With a telephony transmitter conditions are different, in that the amplitude of the oscillation is continually varying in accordance with the modulation, and this variation, which is introduced at an early stage of the proceedings, must be faithfully followed in the subsequent amplifying stages. Hence Class B operation is always used, since, as already explained, the anode power with a Class B valve is controlled by the grid swing, and is therefore proportional to the input.

One of the great advantages of the drive circuit is the fact that it is possible to modulate the oscillation at a point where the power is only a fraction of the final value. As we shall see in the next chapter, modulation usually requires a power four or five times as great as the oscillating power, so that we are able to effect considerable economy by this means.

The modulation is also improved by being applied to the amplifying stage and not direct to the oscillator, since in the latter case there is liable to be some variation in frequency as well as amplitude, and this is most undesirable. The subject, however, is discussed in greater detail in the next chapter.

Constancy of Frequency.

It is most important that the frequency of the oscillations shall not vary. The carrier frequency with a normal short-wave transmission may be anything up to 10 Mc/s. so that a change of 1 per cent is 100 000 c/s., which is about ten times as great as the maximum modulation frequency. If we design our receiving equipment to accept the carrier frequency and a band of 10 000 cycles on each side, it will be clear that a variation of less than 1 per cent on the transmitter would still be sufficient to throw the carrier right outside the acceptance band of the receiver. Even an accuracy of 0·1 per cent is not good enough, and the frequency has to be kept constant to 0·01 per cent at least, and preferably even closer than this.

Now there are a great many factors which will cause variations of this order. Small changes in the voltage, either on the filaments or the h.t. supply to the valves, produce a change in the effective anode impedance, and since this is shunted across the tuned circuit an appreciable change in frequency results.

Attempts have been made to minimize this form of variation by specially designed circuits. If either the grid or anode impedance of the valve can be kept constant, the variation is considerably reduced. One method of doing this is to use a high value of grid leak on the oscillator. Under these conditions the grid resistance of the valve is approximately half that of the grid leak, and therefore tends to remain constant.

The grid leak, however, cannot be made too high, or a phenomenon known as *grid tick* occurs. If the grid voltage momentarily becomes more negative, the oscillations will fall in amplitude. This change must immediately be accompanied by a fall in grid voltage if stability is to be maintained. If the time constant of the grid circuit is too high, the grid voltage will not fall rapidly enough, and the falling off of the oscillations will continue until oscillation ceases altogether. In due course the grid condenser will discharge, and when it does the oscillation will re-commence, and this process will continue indefinitely, the circuit falling in and out of oscillation with a rapid ticking sound.

More elaborate methods for maintaining constant frequency

have been devised, based on the fact that any change in the frequency of the current supplied to a tuned circuit is accompanied by a rapid change of phase, particularly with a good circuit having a high Q.* This in itself tends to correct the frequency, so that a partial protection results. Circuits are possible, however, which contain an impedance in the anode or grid circuit so related to the remainder of the circuit that any change in the phase of the oscillating current automatically corrects the frequency.

The reader should refer to an article by A. P. Llewellyn entitled, "Constant Frequency Oscillators" (*Proc. I.R.E.*, Dec., 1931).

There still remains the variation resulting from changes in the dimensions of the coils and the condensers used, due to atmospheric conditions, changes of temperature, etc., and numerous elaborate and ingenious arrangements have been devised to overcome this. In a condenser the factors affecting the capacitance are the area of the plates and the spacing between them. A rise in temperature will produce an increase in the area due to expansion and also a decrease in the spacing, both of which produce an increase in capacitance. One method of overcoming this is to use a material for spacing the plates which expands more readily than the material of the plates themselves, so that the spacing between the plates actually increases with increase in temperature. If this is correctly designed, it can compensate for the changes in capacitance produced by other causes. It will be clear that there are other ways of tackling the problem, which are too numerous to be discussed individually.

The inductance of the coils is also affected by changes in temperature, and similar methods can be adopted by arranging that the former on which the coil is wound expands axially at a rate which counteracts the change in inductance due to the varying diameter, since this will produce a reduction in spacing between the turns which will tend to reduce the inductance.

Another method which has been used is to turn grooves in a quartz former and to deposit a thin layer of copper in the

* The symbol Q denotes $L\omega/R$.

bottom of these grooves. Since the expansion coefficient of quartz is quite small, this gives a high degree of constancy.

Another method which may be used as an alternative, or for extreme accuracy, as well as previous methods, consists in mounting the whole of the oscillator circuit in the constant temperature oven previously mentioned, and so important is the question of frequency stability that elaborate precautions such as this are quite often adopted.

Crystal Control

Modern conditions, however, are so stringent that even these

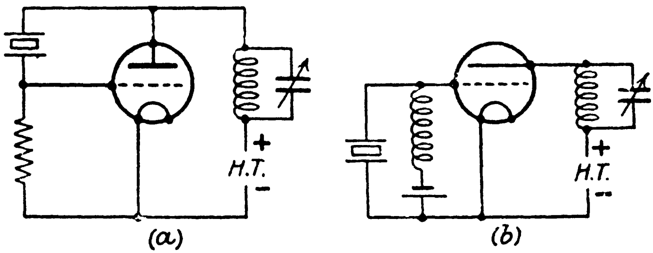

FIG. 47. CRYSTAL-CONTROLLED OSCILLATOR CIRCUITS

precautions are only partially successful, and recourse is had to crystal control, which depends upon what is known as the *piezo-electric* effect. When a crystal of quartz is cut in a certain direction relative to the axis, it is found that an e.m.f. applied across a pair of plates, one on each side of the crystal, causes a mechanical expansion and contraction in a direction at right angles to the electric field. This movement is normally very small, but if the applied voltage is alternating in character, and the frequency is varied, a resonance effect is obtained, there being one particular point at which the movement becomes very large and appreciable energy is absorbed.

Now this resonance is exceedingly sharp and occurs over a frequency variation of a few parts in a million, so that it obviously forms an ideal control for an oscillator, particularly as the range of frequency covered by quartz crystals of a reasonably manageable size falls conveniently within the range we require.

Fig. 47 illustrates two simple types of circuit, the more usual

being that shown at (*b*). The crystal replaces the customary tuned circuit, so that in diagram (*a*) the arrangement is similar to an ultraudion circuit, while diagram (*b*) is equivalent to a tuned-grid circuit.

The anode circuit of the valve is roughly tuned in each case, but this tune is not critical. The tune is adjusted to a frequency slightly higher than that of the crystal, so that the input resistance of the valve is negative (due to Miller effect). This offsets the losses in the crystal and continuous oscillation results. The variation of setting of the anode tune alters the waveform of the oscillation to some extent, and also the amplitude, and the setting is therefore adjusted in practice to give the best results.

In Fig. 47 (*a*) the grid is tied to the cathode through a leak, while in the diagram (*b*) an h.f. choke is used. Either method may be used with either circuit. A simple leak is less expensive, but since the crystal acts as a small condenser it acquires a negative charge due to the grid current which flows on the peaks and therefore runs the valve back to a condition of comparatively low slope. With an h.f. choke leak this action does not occur, and therefore the valve must be provided with bias to bring it to a suitable operating point. This is shown as a battery in the diagram, though the customary cathode bias may be adopted if required. In general, a circuit with a choke leak will give more output than with the resistance leak.

The actual frequency at which the crystal operates depends upon the manner in which the slice is cut from the whole crystal. Fig. 48 shows the general form of a quartz crystal, hexagonal-sided with irregular ends. The vertical or optical axis is known as the Z axis, any of the three axes across the corners of the hexagon are known as X axes, and the axes at right angles to these are known as Y axes.

The two most common types of slice are those cut perpendicular to one or other of these axes. An X-cut crystal is one which has its face perpendicular to an X axis, while a Y-cut crystal has its face perpendicular to a Y axis.

X-cut crystals will exhibit two modes of oscillation, one corresponding to the thickness (i.e. parallel to the X axis) and the other corresponding to the width parallel to the Y axis. In both cases the frequency is given by Terman as $2\,860/d$ kc/s.,

where d is the thickness (t) or width (w) in mm. Ladner and Stoner give a slightly different figure of 2 730/d.

X-cut crystals can be supplied to oscillate over a range of 250 to 10 000 kc/s., and exhibit a negative temperature coefficient of the order of 20 parts per million per degree centigrade.

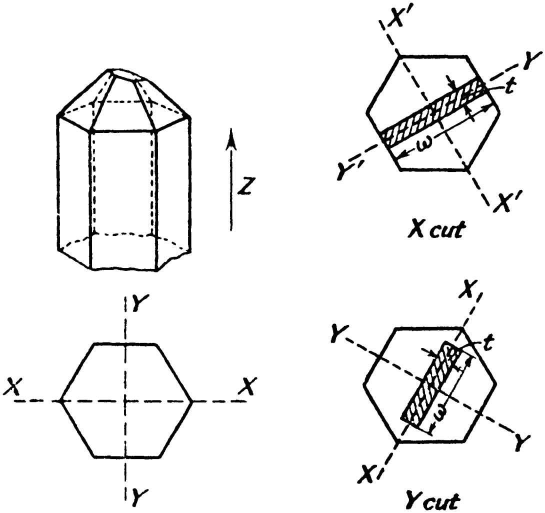

FIG. 48. DIAGRAM OF QUARTZ CRYSTAL SHOWING PRINCIPAL AXES

Y-cut crystals will operate again in two modes, but the more satisfactory one is that corresponding to the width parallel to the X axis. The frequency of oscillation for this is given by the same formula as for X-cut crystals.

The other mode of vibration is inclined to be erratic and depends upon the ratio of thickness to width, but as long as this ratio is small the frequency is given approximately by 2 000/t kc./s., where t is the thickness in mm. as before. This form of crystal exhibits a positive temperature coefficient of 35 to 40 parts per million per degree centigrade.

Effect of Temperature.

The fact that one form of cut will give a negative tempera-

ture coefficient while the other will exhibit the opposite form of variation obviously suggests the possibility of combining the two to obtain a crystal having a zero temperature coefficient. This is now done in special cases, and slices using

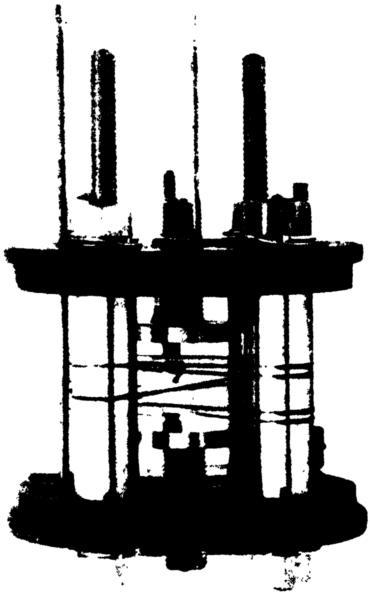

FIG. 49. QUARTZ CRYSTAL IN TEMPERATURE-CONTROLLED HOUSING

(*Marconi's Wireless Telegraph Co.*)

what is called an *A.T.* or *constant-temperature* cut are available having a very small change in frequency with temperature. Such a crystal of course is of considerable value, and is likely to be increasingly adopted.

In the absence of such a cut, however, the ordinary *X* or *Y* cut may be used in a rigid holder such as that illustrated in

Fig. 49, the whole being housed in a temperature-controlled chamber. These temperature-controlled ovens are very often double, containing an inner and an outer oven, each temperature controlled, the inner being at a slightly higher temperature than the outer, and the outer being kept at a temperature higher than the highest ambient temperature ever likely to be encountered in the particular climate.

Crystal Holders.

It is often considered that a crystal will generate a constant frequency irrespective of the conditions of use. This is not the

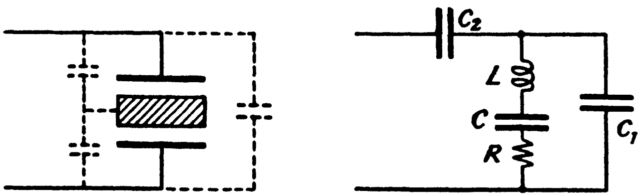

FIG 50. EQUIVALENT CIRCUIT OF CRYSTAL

case, however, for the holder in which the crystal is mounted has some effect. Fig. 50 illustrates the equivalent circuit of a crystal. The crystal being a mechanical resonant circuit has mechanical inertia and elasticity, which can be represented on the equivalent circuit as inductance and capacitance, while the resistance R represents the loss in the crystal due to molecular friction.

It will be clear that there are two capacitances which are variable and which can affect the frequency. One is the capacitance C_1 across the crystal, and this is sometimes used to provide a small measure of variation so that a crystal may be pulled in accurately to some particular frequency. This condenser, however, cannot be made too large, or the crystal will cease to oscillate.

The second effect is that of a series capacitance between the metal plates of the holder of the crystal itself, represented by C_2. This again can be used to pull the crystal slightly, provided the variation required is not large. As a general rule, the series condenser method is used for low-frequency crystals up to a

few megacycles and the parallel method for frequencies higher than this.

The possibility of varying the frequency in this manner, however, makes it clear that precautions must be taken if absolute constancy is required to make sure that the mechanical dimensions of the holder do not change appreciably with time or temperature. The order of variation is quite small, but the accuracy of crystals is so high that the crystal itself may easily be better than its holder in respect of constancy.

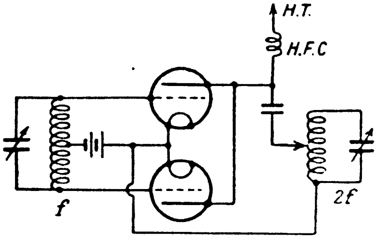

FIG. 51. FREQUENCY DOUBLING CIRCUIT

It may also be noted that grease, wax, or similar substances on the face of the crystal may prevent it from oscillating, while dust or scratches will cause erratic oscillation. Crystals should therefore be handled with extreme care and are usually mounted in sealed holders to prevent damage by mishandling.

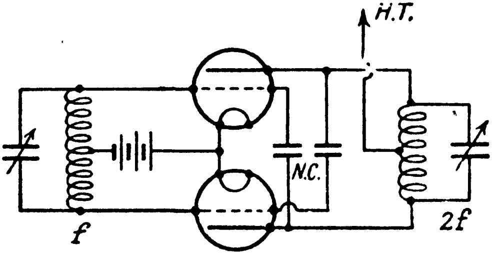

FIG. 52. NEUTRALIZED FREQUENCY DOUBLER

Frequency Doubling.

With high frequencies the thickness of the quartz slice becomes too small for practical conditions, and it is customary therefore to generate a frequency which is a sub-multiple of the value required. This frequency is then doubled once or twice by suitable circuits. The most usual is a variant of the push-pull arrangement, and is shown in Fig. 51. The input

circuit is tuned to the fundamental frequency and the anode circuit to twice this frequency. Since one of the anodes receives an impulse every half-wave, which occurs at twice the fundamental frequency, it will be seen that the anode circuit is satisfactorily energized. The valves, of course, are biassed to cut off.

This circuit, however, is unsymmetrical, and the arrangement of Fig. 52 is often used for short-wave work. This is

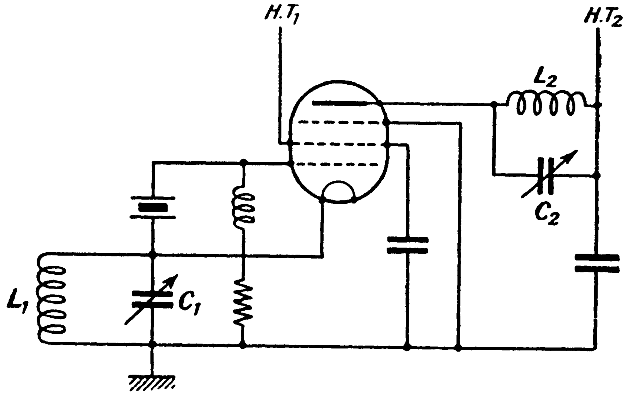

FIG. 53. TRITET OSCILLATOR

L_2C_2 is tuned to twice the crystal frequency.

symmetrical and neutralized. It is, however, necessary to extinguish one of the cathodes, so that only one valve is working, and the efficiency therefore is low.

Two or even three stages of frequency doubling are sometimes used, and the symmetrical push-pull arrangement shown is preferred for the reasons already given, i.e. maintenance of stability and the avoidance of parasitic oscillations.

Tritet Oscillator.

A combined crystal oscillator and frequency multiplier has been devised by Lamb and is illustrated in Fig. 53. This is known as the *tritet oscillator*, since it uses a triode oscillator in conjunction with a tetrode or pentode frequency multiplier. An ordinary tetrode or pentode is used, the first two electrodes being operated as a triode to maintain the oscillation in the

crystal, and the main anode circuit being tuned to twice or three times the frequency. This circuit is utilized for low-power transmitting stations, high-power commercial stations using separate stages for the generation of the oscillation and for the multiplication.

Mechanical Oscillators.

The quartz crystal is not the only form of oscillator which can be used, and it has certain disadvantages which have prompted engineers to seek other solutions. A simple steel rod can be made to oscillate longitudinally in frequencies of the order of 20 000 c/s. without difficulty, and the frequency of the oscillation is almost entirely dependent upon temperature, so that by enclosing the oscillator in a suitably-controlled oven very constant oscillations can be developed.

The frequency, of course, is much too low for short-wave working. It would have to be doubled up by successive stages. Eight stages of doubling, however, would bring us into the 80-metre band and, although this may seem a very cumbersome arrangement, it has actually been used.

The output from the frequency doubling stages is fed into the amplifier in the normal way, modulation or key control being introduced at a suitable point as required.

Feeder Stabilization.

This review of the subject is necessarily brief, and the reader who wishes to investigate the matter further should refer to "Methods of Frequency Control," by Conklin, Finch and Hansell, *Proc. I.R.E.*, Nov., 1931; and to A. S. Angwin's "Chairman's Address," *Journal I.E.E.*, Dec., 1931. Mention may be made in conclusion of a method which has some application where a smaller degree of constancy is tolerable. Here the regeneration from anode to grid necessary to sustain oscillation is obtained through a concentric feeder so adjusted and terminated that the phase at the grid end is just correct. Any variation in frequency causes a shift in phase at the grid, which is in the correct sense to alter the frequency and bring the oscillator back to normal working again.

The circuit has the advantage that it can be applied to an oscillator of considerable power, whereas all the methods so far described are only suitable for oscillators delivering

very feeble power and operating practically in an unloaded condition.

A modification of the scheme specially suitable for ultra-short waves is discussed in Chapter IX.

Wavemeters.

The measurement of frequency (or wavelength) is of considerable importance. For this purpose a wavemeter is used, consisting of a tuned circuit, so constructed as to minimize any variation of the inductance or capacitance with atmospheric changes, loosely coupled to a suitable detecting device. For crude estimations a flashlamp may be included in the circuit. This will glow when the current rises high enough, and by locating the wavemeter near the oscillating circuit under test it can be arranged that the lamp just lights at one point on the scale, when the wavemeter is in resonance.

This is not accurate enough for modern usage, however, and the usual practice is to couple the circuit very loosely to a valve detector. The coupling must be very weak so that it introduces negligible load on to the circuit. Otherwise the constancy of calibration may be impaired. In addition, some special device is adopted to locate the resonance point exactly. For example, a small fixed condenser may be connected with a key switch across the main condenser and the wavemeter adjusted till the closing of this switch makes no difference to the detector reading. Then with the switch up, the wavemeter is above resonance, while with the key depressed it is below resonance by an equal amount, and the true resonant point midway between the two can be determined much more exactly than by trying to tune to the somewhat broad peak.

In a crystal oscillator the frequency is positively determined by the crystal which is pre-calibrated. The actual calibration of crystals and wavemeters to the close limits demanded by modern practice is a highly specialized subject. The reader should refer to "A Self-contained Harmonic Wavemeter," by Dye, *Phil. Trans.*, Nov., 1924, and "Quartz Resonators and Oscillators," by Vigoreux, H.M. Stationery Office, 1931.

CHAPTER VII

MODULATION

A CONTINUOUS carrier wave will not communicate any intelligence. It must be modulated in some way, either by being broken up into dots and dashes or by the superposition of speech. Both these effects are usually produced at an early stage of the proceedings, where the power handled is only a fraction of the total radiated power. Keying, for example, can be obtained either by stopping the master oscillator from functioning, or preferably by rendering one of the early valves in the amplifier inoperative A convenient way which is often used is to over-bias one of the amplifier valves, this bias being reduced to normal when the key is depressed, and it is not necessary to discuss the methods in great detail. The circuit adopted, of course, must be instantaneous in character, so that the oscillator current rises very rapidly to its full value and cuts off equally sharply, thus making high-speed keying possible.

In telephony the amplitude of the oscillations has to be varied at an audio frequency, and for this purpose it is customary to control the amplification of one of the early valves in the drive circuit. As already explained, these drive valves are operated on Class B principles, so that if the amplitude at the beginning of the chain is varying, these fluctuations are faithfully transferred to the output.

Direct modulation of the oscillator is possible, but is rarely used on short waves because the variation of amplitude is usually accompanied by a small change in frequency, and this cannot be permitted. By modulating in the amplifier stages, the frequency is clearly unaffected.

Choke Modulation.

The most common form of modulation is the choke control method, often known as the *Heising* system. In this, the valve to be modulated is fed in parallel with a modulator

valve through a low-frequency choke, as shown in Fig. 54. The effective anode voltage on the two valves is thus the h.t. voltage plus or minus any audio-frequency voltage developed in the choke, so that if an audio voltage is applied to the grid of the modulator valve the anode voltage of both valves will fluctuate accordingly. Now if the amplification of the radio frequency valve is made proportional to the anode voltage, then the amplitude of the radio-frequency currents will be

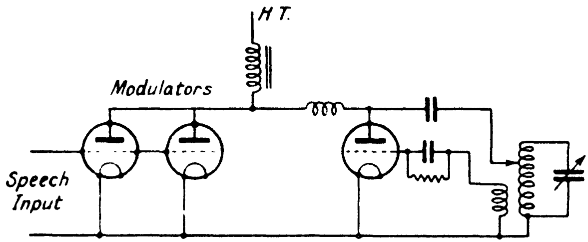

FIG. 54. CHOKE-CONTROLLED MODULATION CIRCUIT

controlled in conformity with the modulation impressed on the grid of the modulator, and this method is both simple and effective in use.

As a result of this the amplitude of the carrier wave varies at a frequency determined by the applied modulation and to an extent controlled by the conditions. Let us consider for a moment modulation of a single frequency. The modulated wave will appear somewhat as shown in Fig. 55. The top line shows a partial modulation and the second line shows full modulation where the carrier has just been reduced to zero, and the third line shows over-modulation where the carrier is reduced to zero too soon. This latter condition obviously introduces distortion and must be avoided, the maximum possible modulation being that which just reduces the carrier to zero so that the amplitude swings between nothing and twice the normal value. In practice, the maximum modulation used is rather less than this, since the curvature of the valve characteristics towards the zero current axis tends to introduce distortion. Moreover, the detector used at the receiver will usually only handle a modulation of some

80 per cent without distortion, as explained in *Modern Radio Communication*, Vol. II.

Depth of Modulation.

The ratio between the amplitude of the modulation wave and the carrier wave is called the *depth of modulation*, and

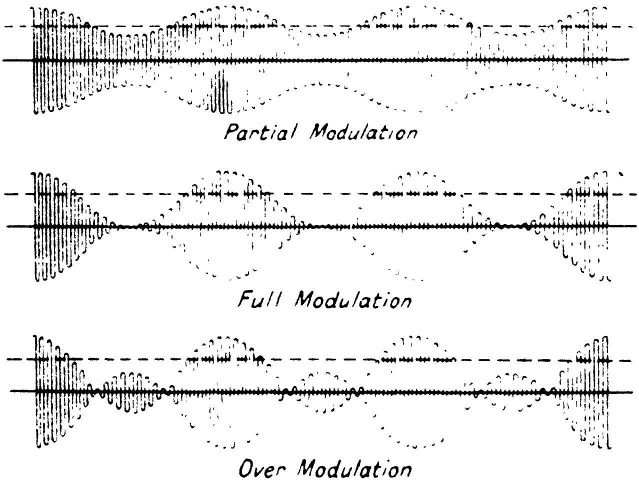

Fig. 55. Diagram of Modulated Carrier Wave

obviously if the carrier is just reduced to zero, these two are equal, giving us a modulation of unity or 100 per cent. In the top line of Fig. 55 we have 50 per cent modulation, the carrier being reduced to half its normal height by the modulation.

Mathematically, a modulated carrier can be expressed in the form $e = E \sin \omega t\,(1 + m \sin pt)$, where ω is the (angular) carrier frequency, p the modulation frequency, and m is the depth of modulation.

Now it will be clear that a carrier wave modulated 100 per cent varies in amplitude between zero and twice the normal value. Consequently the peak power, being proportional to the square of the current, is four times the normal. This extra

power must be supplied from somewhere, and actually comes from the modulator valve, which therefore has to be of appreciably greater capacity than the valve which it is controlling. Actually in practice the modulator valve has to dissipate three to four times as much power as the oscillator (or amplifier) valve, the position being somewhat aggravated by the fact that the modulator has to operate in a Class A condition with a fairly large mean anode current.

Other Forms of Modulation.

The comparative inefficiency of the choke control system has naturally led to attempts to find other methods, but no really satisfactory alternative has been evolved. Modulation by the control of the grid voltage is sometimes used, since the amplitude of the oscillations with a Class B amplifier can be made proportional to the voltage on the grid, within certain limits. The control, however, begins to lose effect as the depth of modulation increases (i.e. as we try to reduce the carrier current to zero), and hence some distortion is liable to arise with high percentage modulation.

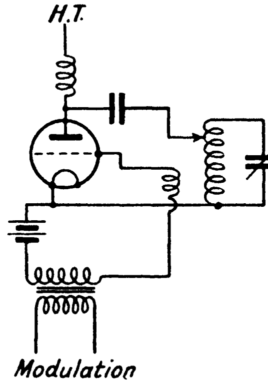

FIG. 56. SIMPLE GRID-MODULATED TRANSMITTER

On the other hand, the method does not consume any appreciable energy and is more sensitive than the choke case, so that for small powers it has much to commend it. Fig. 56 shows a grid-modulated circuit applied to a simple transmitter. An alternative method is to use a valve as the grid leak in a circuit of the type shown in Fig. 40.

Still another method is the series circuit shown in Fig. 57, where the h.t. supply to the oscillator or amplifier is fed through the modulator valve. If under normal conditions the two valves take approximately the same current, it will be clear that the voltage will divide equally over the two. The application of voltages to the grid of the modulator will cause the internal resistance to vary, and hence a greater or less

proportion of the voltage to be developed across the oscillator, thus modulating the oscillating current. The oscillator or amplifier being modulated must clearly be operating in a Class A condition with this circuit, and there is also a limit to the extent to which the modulation can be carried out, for the change in internal resistance of the modulator valve only remains proportional to the input voltage over a relatively small region, and hence distortion will result at the higher modulation depths. On the other hand, the method has the advantage that it uses no impedances and therefore its frequency characteristic is good. It will modulate down to zero frequency and its upper limit is only set by the shunt capacitance effects, which can, with proper precautions, be reduced to quite small limits.

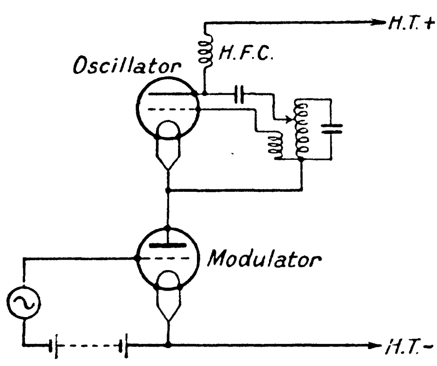

FIG. 57. SERIES MODULATION CIRCUIT

Side-band Theory.

It now becomes desirable to consider modulation from a slightly more mathematical viewpoint and to introduce the conception of side-bands.

Often in alternating-current practice it is convenient to regard some phenomena as being the result of certain imaginary effects, and our ideas of modulation are helped by a similar fiction. When we modulate a carrier wave with a pure tone, we produce a complex wave of which the amplitude is varying, and with our knowledge of the objection which is always raised by Nature to any deviation from its set course, it will be understood that the behaviour of the circuit when presented with a wave form of this nature is complex. Now actually we find that we can produce the same effect as that of modulation by transmitting three different frequencies

at the same time. One of these frequencies is the original carrier frequency, and the other two are what we call *side frequencies*, which differ from the carrier by the modulation frequency.

For example, suppose we have a carrier wave of 1 000 kc/s. and we modulate this with a frequency of 1 000 c/s., causing the amplitude to vary a thousand times a second, we shall produce a modulated wave similar to that shown in Fig. 55. But we could produce exactly the same wave by transmitting a steady carrier of 1 000 kc/s., and two additional carrier waves having frequencies of 1 001 and 999 kc/s., and it is often helpful to consider modulation in this light. We assume that the presence of modulation has not affected the original carrier, but has introduced two side frequencies 1 000 c/s. above and below the original carrier. The amplitude of these side frequencies would depend on the depth of modulation, and if we have a complex modulation involving a number of frequencies, we say that each one introduces its appropriate side frequencies, giving us what we call *side-bands* on either side of the carrier.

Reception of Modulated Wave.

Support for this viewpoint is obtained from the behaviour of tuning circuits set to receive a modulated wave. If we have tuned to the carrier a circuit having a which reduces the strength of the current to half the maximum value 1 kc/s. off the resonant point, we then find that if we apply to this circuit a carrier wave modulated with a 1 kc/s. note, the depth of modulation which we receive is only half what it should be.

Now if we regard our modulated wave as being due to a carrier and two side frequencies 1 kc/s. off tune, the reason for this will be clear. The carrier wave will, of course, be received at its full strength, since the circuit is tuned to this frequency, but since the resonance curve is such that frequencies 1 kc/s. off tune are only received at half strength, the two side frequencies will only produce half their correct response, so that the effective modulation of the received signal will only be half what it should be.

In short, we can regard a modulated signal as being made up of a band of frequencies spreading out on either side of the

carrier to an extent corresponding with the maximum modulation frequency. Mathematically, we can show the existence of these side frequencies by simple trigonometrical transformations working from the expression for modulation given previously.

Let the modulated wave be represented by

$$e = E \sin \omega t\,(1 + m \sin pt).$$

We can expand this into the form

$$\begin{aligned} e &= E \sin \omega t + mE \sin \omega t \sin pt \\ &= E \sin \omega t + (mE/2) \cos (\omega - p)t - (mE/2) \cos (\omega + p)t, \end{aligned}$$

which is the original carrier wave with two side frequencies, on each side, of amplitude $m/2$ times as great. Hence for 100 per cent modulation ($m = 1$) the side-frequency amplitude is one-half that of the carrier.

It must be clearly understood that there is no question of this side-band theory being more correct than the original idea of varying amplitude. It is merely another way of regarding the same phenomenon. Many of the effects obtained with short-wave working become more readily understandable when viewed in this way, one specific example being the distortion and mutilation of signals due to reflection at the Heaviside layer. The amount of bending produced in the upper atmosphere is proportional to the frequency of the wave, and in conditions where the boundaries of the layer are changing quite a small difference in frequency may have a large effect.

In particular, the rotation or twisting of the plane of polarization, discussed in Chapter II, will vary with the frequency, so that we may find the carrier and side-bands differently polarized after reflection. In an extreme case, the plane of polarization of the upper side-bands might be twisted 90° relative to the carrier. The ordinary receiver, of course, only responds to vertically or horizontally polarized waves, but not both. Our received wave will, at one instant, have a vertically polarized carrier and horizontally polarized upper side-bands. Hence the receiver will deliver full amplitude on the carrier wave, but nothing from the upper side-bands and only partial reception from the intervening frequencies. As the plane of

polarization gradually rotates, we shall obtain maximum response from the upper side-bands and nothing from the carrier. This unequal reception produces distortion. The loss of the upper side frequencies means that the upper tones of our speech or music are cut off and the clearness of articulation is lost, while the reception of the upper side-bands only, without the carrier, produces a meaningless jabber.

Various attempts have been made to overcome this difficulty, which is called *selective fading*, as opposed to the ordinary amplitude fade, where the whole signal varies in strength at a relatively slow rate. Both effects, of course, are produced by uneven reflection in the upper atmosphere, but whereas an amplitude fade can be compensated to some extent by the use of automatic gain control, selective fading cannot be treated in the same way. One solution is to use what is called *diversity reception*, and for this purpose a series of receivers is used spaced from one another by a few wavelengths. These aerials feed individual receivers, and the outputs are mixed and fed into the audio-frequency amplifier. The subject has already been discussed in Chapter V.

Suppressed-carrier Working.

The mathematical analysis of a simple modulated wave shows that of the total power, 50 per cent is supplied to the carrier and 25 per cent to each of the side-bands. Now it is only the side-bands which are effective in communicating the information required, and actually only one of the side-bands is necessary, so that we are wasting 75 per cent of the power. It is not surprising therefore that many attempts have been made to dispose of the carrier altogether. The first move in this direction was known as *suppressed carrier* working, in which, having first modulated the carrier and produced the side-bands, the carrier frequency was removed. Since our carrier frequency is unaltered by the modulation and remains constant, both in frequency and amplitude, it is quite possible to do this. The side-bands were then transmitted, and at the receiving end the carrier frequency was re-introduced by means of a local oscillator adjusted to the correct frequency and amplitude.

Unfortunately, there are certain practical difficulties which

render the scheme quite unsatisfactory. In the first place, the frequency must be very exact, as otherwise the two side frequencies caused by a given modulation frequency will not recombine correctly, but will give us two notes. For example, a 1 000 cycle note on a 1 000 kc/s. carrier would give us side frequencies of 999 and 1 001 kc/s. If these side frequencies at the receiving end are mixed with the carrier of 1 000·1 kc/s. we shall obtain modulation frequencies of 900 and 1 100

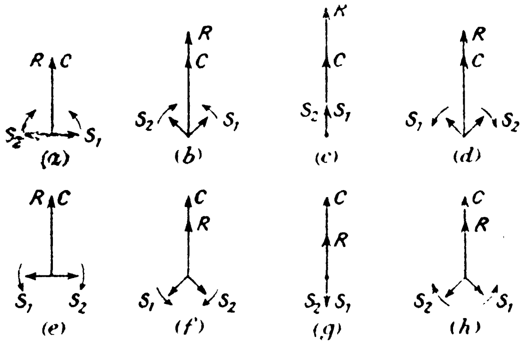

FIG. 58. VECTOR DIAGRAM OF MODULATED CARRIER WITH SIDE-BANDS IN PHASE

cycles, i.e. two different notes instead of the single 1 000 cycle note which we should have.

Phase of Modulation

An even more serious difficulty, however, is that of phase. The frequency could be kept sufficiently constant by the use of crystal-controlled circuits, but it is found that the phase of the modulation has a most important effect on the results, and before discussing the matter further it would be as well to consider this phenomenon.

Briefly, the effect is as follows. If we apply modulation to a carrier wave, we shall produce a variation in the amplitude in the manner already described, but if we now alter the phase of the carrier, we shall find that the variations in amplitude are decreased, and, in fact, with deep modulation and a 90° phase shift, it is possible for the modulation to disappear altogether.

A particularly simple way of regarding the phenomenon

has been put forward by Ladner and Stoner, who have treated the matter in some detail. A modulated wave comprises three frequencies, the carrier and two side-bands, one a little faster and the other a little slower. We can, therefore, regard the carrier as fixed and the two side frequencies as being generated by two vectors of equal amplitude rotating in opposite directions. The resultant amplitude of the side-band vectors and the carrier depends upon the depth of modulation, and Fig. 58

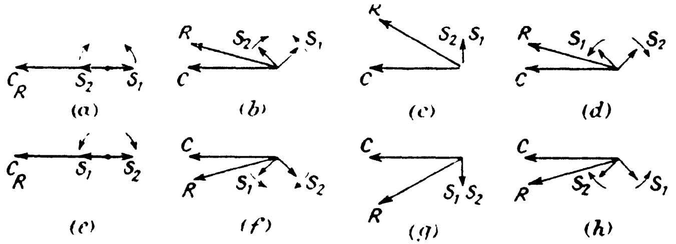

FIG. 59. VECTOR DIAGRAM OF MODULATED CARRIER WITH SIDE-BANDS 90° OUT OF PHASE

shows successive stages of a 25 per cent modulation cycle where the side-bands start off in phase with the carrier. It will be seen that the resultant vector R, produced by the vector sum of the two side-bands S_1S_2 and the carrier C, varies in amplitude from the maximum to the minimum in a rhythmic manner as we progress through the cycle.

This is the ordinary modulation effect and is quite straightforward, but let us now consider what happens when we displace the actual vector by 90°, as shown in Fig. 59. It will be clear that the variation in the amplitude of the resultant wave is now nothing like as great as before and, in point of fact, if the modulation were 100 per cent, we should obtain practically no variation at all, but merely a swinging of the resultant vector.

It will be clear, therefore, that any suppressed carrier system depends for its success upon the re-introduction of the carrier *in the exact phase which it occupied at the transmitter*, and this is a physical impossibility with our present knowledge of the art. The system can be used with telegraphic communication where the phase of the modulation is of minor

importance, but generally the system has given way to the modification known as *single side-band* working, which is just as efficient and overcomes many of the difficulties.

Single Side-band Working.

In this system, both carrier and one side-band are suppressed, leaving the remaining side-band to be transmitted to the receiver. The carrier is re-introduced, and since we have only one side-band, neither the frequency nor the phase of the carrier require to be exact. Reference to the diagrams just shown will make it clear that where we are only dealing with one side-band, the question of phase does not produce any alteration in the form of carrier variation, but merely a small difference in the time at which the events occur.

It is generally assumed that the re-introduction of the carrier converts the single side-band into the original wave, but this is not right. It gives quite a close approximation, but actually the amplitude variation is not quite the same as in the original wave, so that some distortion is introduced, while there may be quite considerable phase-displacement of the various modulation frequencies relative to one another.

Now for aural reception, this is not important, because the ear is unable to distinguish phase differences in the various tones which go to make up a composite sound except in the case of transient sounds, and the amplitude distortion can be minimized by using a large carrier at the receiving end. Hence for speech transmission this single side-band arrangement is very successful and is used to a considerable extent.

Increasing use is being made of facsimile transmission, in which the question of phase is of the greatest importance, and for this the single side-band system is not very successful, so that the straightforward transmission by carrier with both side-bands has to be used.

Carrier Removal Circuit.

The removal of the carrier wave is usually effected by a sort of push-pull arrangement, as shown in Fig. 60. Just as a push-pull circuit balances out the steady d.c. in the anode circuit, so the Carson circuit balances out the carrier wave. The carrier is introduced to both grids in the same phase,

whereas the modulation is applied in the normal push-pull manner, so that the grids are in opposite phase. The side frequencies, therefore, produced by interaction between the modulation and the carrier add up in the output circuit, whereas the carrier is balanced out and we are left with the side frequencies only.

The removal of the unwanted side-band is now a simple matter of filtering; the output being passed through a

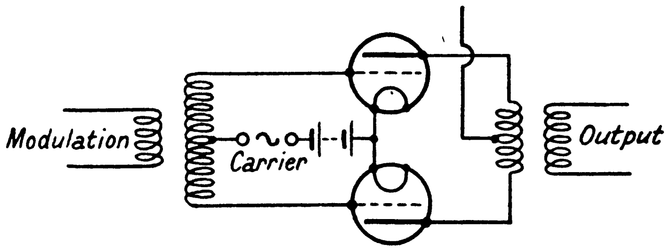

FIG. 60. CARSON BALANCED MODULATOR CIRCUITS

band pass filter which accepts one side-band and rejects the other.

Eckersley has suggested that for telephony transmission the best results are obtained if a small percentage of the carrier is still left. For a detailed discussion of the subject, the reader should refer to "Asymmetric Side-band Broadcast Transmissions," *Journal I.E.E.*, Vol. 77, p. 517.

Telegraph Band Width.

The question of the frequency spread or *band width* of a transmission has so far been considered in relation to a telephony transmission. It is clear, however, that some side-bands are necessary even for the transmission of Morse signals. It is impossible to make or break a circuit instantly, the current taking appreciable time to rise or fall. To preserve a reasonable wave shape for the dots and dashes, the rise and fall should not occupy more than, say, one-tenth of the duration of a dot.

At normal keying speeds a dot lasts about $\frac{1}{5}$ sec., so that the rise or fall of the current must take place in 1/50 sec.

This corresponds to one-half of a cycle viewed from an a.c. standpoint, so that our effective modulation frequency is 25 cycles per sec.

High-speed transmission may involve speeds ten or fifteen times greater than this, but, even so, the modulation frequency only rises to between 200 and 400 per sec., so that a pass-band at the receiver of 0·5 to 1 kc. is adequate.

CHAPTER VIII

SHORT-WAVE RECEIVERS

THE short-wave receiver varies in its construction according to the requirements. For simple reception, principally of broadcast or amateur telephony transmissions, the circuits and general arrangement can be of a very modest character, while for commercial services much more elaborate equipment

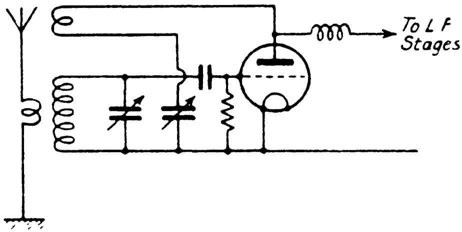

FIG. 61. SIMPLE REGENERATIVE RECEIVER

is used in order to ensure continuity of service as far as possible. In every case the aerial system has an important bearing on the results obtained, but this aspect of the question has already been discussed in Chapter V, and will not be referred to in any detail in the present chapter.

The simplest form of circuit is the straightforward regenerative receiver shown in Fig. 61. Here the aerial is coupled to a tuned secondary circuit, which is applied to a suitable detector, usually a grid rectifier. Reaction is used from the anode of the valve in order to improve the sensitivity, and telephones may be used for the reception. More usually the valve feeds a low-frequency amplifier, so that an additional amplification is obtained sufficient to operate a loud speaker.

Obtaining Smooth Reaction.

Now the construction of a circuit of this type is straightforward and follows the usual practice, except that allowance

must be made for the high frequencies involved. Moreover, in order to obtain really good sensitivity, the reaction circuit must be particularly smooth. By this is meant that an increase in the regeneration from the anode circuit must cause a gradual improvement in the signal strength until a point is reached where the circuit gradually slides into oscillation. Many circuits do not conform to this desirable requirement. When they reach the oscillating point they suddenly commence to oscillate violently with a sharp click or plop in the loud speaker, and with a circuit in this condition it is not possible to work at the threshold of oscillation.

Another effect known as *backlash* often occurs, this being the condition when the reaction setting required to stop oscillation is appreciably less than that necessary to start it, so that when the receiver commences to oscillate a reduction in the reaction setting does not immediately check the oscillation, and by the time the oscillation has been stopped, so that the receiver reverts to its non-oscillating condition, it is nowhere near its best sensitivity. For really satisfactory working, the receiver must be capable of being brought right up to the point of oscillation, when it will slide almost imperceptibly into the oscillating condition, and at any point a reduction in the reaction setting immediately causes the set to revert to its previous condition. This is known as *working on the threshold of reaction*, and is obtained by proper attention to the working point on the characteristic.

In general, backlash trouble is usually found to be due to the manner in which the anode current varies. With a grid detector an increase in the signal is accompanied by a reduction in the anode current. This causes a reduced voltage drop in the anode circuit, an increase in the anode voltage, and hence an effect tending to increase the reaction. This produces further decrease in anode current and the effect is cumulative, so that the circuit "runs away" and commences to oscillate with the click already mentioned. Clearly a small reduction in reaction setting will have little power to check the oscillation, and quite appreciable loosening of the reaction setting is necessary.

To avoid this effect we must work the valve in such a condition that small changes of h.t. voltage do not appreciably

alter the amplification. This can be achieved by suitable choice of valve, h.t. voltage, and grid condenser-leak combination.

Threshold howl is another effect caused by a reverse process. If the valve is so operating that an increase of signal causes a reduction in sensitivity, due to a fall in anode voltage or similar cause, the commencement of oscillation will be accompanied by an automatic reduction of the reaction which will check the oscillation, and the circuit will fall in and out of oscillation at a rapid rate—usually at an audible frequency, giving rise to the howl referred to.

This defect can sometimes be cured by inserting resistance in the anode circuit, but the basic cure is, of course, the same as before, namely, to operate the valve under conditions where the amplification does not vary with small changes of signal input.

Hand Effect.

It has already been explained that the voltage distribution over a wire which is carrying a high-frequency current is not constant, but varies because of the distributed inductance and capacitance of the wire. Hence points on the chassis of the set, which are supposedly at zero potential, are often not really so, but possess some high-frequency potential. If the hand is brought near these points, currents flow through the body to earth through the capacitance between the operator and the ground, and this seriously affects the tuning.

The difficulty was avoided in the early days by the use of long extension spindles of insulating material, so that it was not necessary for the operator to come within four or five inches of the apparatus, and on some ultra-short wave reception this precaution is still necessary; but a better understanding of the reason for the effect shows that if care is taken to ensure that the chassis (and with it the moving plates of the condensers, etc.) are properly earthed, the difficulty is obviated.

The panel and chassis of the set should not carry any current. All earth return currents should be taken through the wiring, and this wiring connected to the chassis at one point only. Even so, it is possible for currents to be induced in the chassis and panel, and it is sometimes necessary to insulate

the chassis from the panel and again connect the two at one point only, making the panel of heavy gauge aluminium or copper.

Such precautions, however, are usually only necessary in the construction of very special receivers or signal generators, and then only at frequencies of 25 Mc/s. upwards. For most purposes, simple precautions to avoid earth currents in the chassis and panel will be sufficient.

The earth lead should be as short as possible and should be taken direct to a point on the chassis as close as practicable to the tuning condenser. If a long earth lead is essential, then it is desirable to introduce an additional tuning condenser in this lead, so that the standing waves on the earth lead can be adjusted to give a voltage node at the receiver. The same effect could be obtained by accurately adjusting the length of the earth lead, but this is not desirable where reception is required on several different wavelengths. In any case, it is usually only on the shorter wavelengths that any trouble is found, and the practice of tuning the earth lead will usually be found to meet the requirements.

Design of Short-wave Coils.

The coils used for short-wave reception are quite small, the inductances ranging from 1 to 5 μH. in the ordinary way. This inductance is obtained quite easily by the use of a few turns of wire on a small diameter former, and hence the usual construction is of a rather skeleton form. The effect of dielectric loss on short waves is discussed later in the chapter, where it is shown that the former has only a minor influence on the efficiency of the coil. The main source of loss is that of conductor resistance, which is quite appreciable, despite the small number of turns, in comparison with the small inductance. The Q of the average short-wave coil is about 150.

The use of any form of stranded wire is quite impracticable, because the losses in the insulation between the strands more than counterbalance the advantage gained from the stranding. The usual procedure is to use fairly heavy gauge wire with the turns spaced from one another by a distance approximately equal to one diameter. Fig. 62 shows the variation of L and Q for a coil having constant diameter, and the same number

of turns and winding length, but with different gauges of wire, so that the finer wires are relatively farther apart. The broad maximum, when the spacing is roughly equal to the wire diameter, will be seen, but it is clear that this is not critical and the nearest convenient gauge is satisfactory. It is interesting to note that the inductance also varies with the

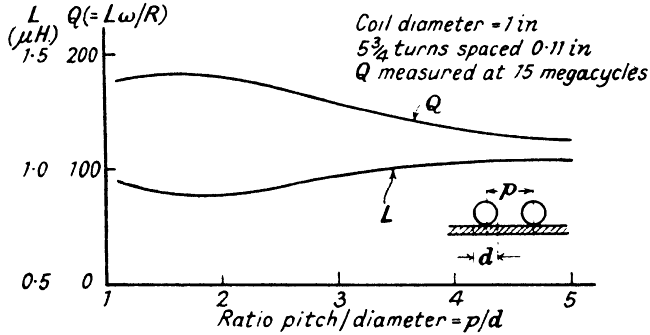

Fig. 62. Variation of Q and L with Spacing of Wire on Coil

spacing to a small extent, and is actually a minimum in the region of maximum Q.

With a spacing of this order and a constant diameter, the Q of the coil increases with the diameter of wire, rapidly at first and then more slowly. The increase of wire size, of course, involves progressively longer coils and in practice, a length equal to the diameter constitutes a good practical rule. As will be seen from Fig. 63, the increase of Q beyond the point is only small.

Fig. 64 shows the effect of coil diameter, again with the wire spaced one diameter (i.e. $p = 2d$), from which it will be seen that with fine wires there is little advantage in a large coil, but with heavier gauges an increased diameter gives an improvement.

The effect of screening is similar to that on longer waves. If the coil is at least half its diameter away from the metal, the reduction in Q is about 10 per cent only, while the inductance drops by about 15 per cent. If the coil is to be screened,

therefore, the inductance should be made a little higher than required to allow for the effect of the screen.

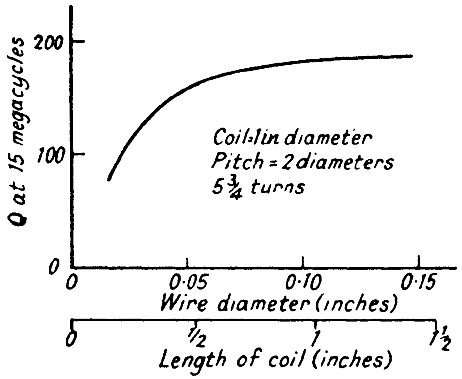

FIG. 63. ILLUSTRATING THAT THERE IS LITTLE ADVANTAGE IN USING A COIL HAVING A LENGTH GREATER THAN THE DIAMETER

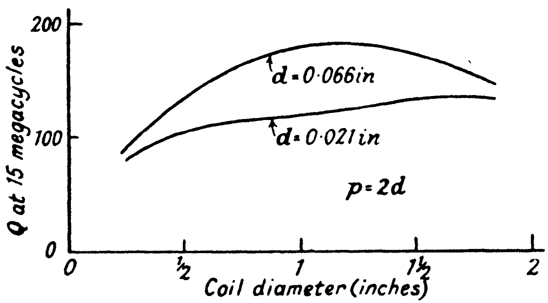

FIG. 64. ILLUSTRATING THE INFLUENCE OF COIL DIAMETER

Inductance Calculations.

The inductance of short-wave coils may be calculated from the usual formulae, and various charts and abacs are available. The author has devised a simple expression which gives a speedy and accurate result, and is easily memorized. If d

and l are the diameter and length of the coil *in inches*, the inductance is given by

$$L = \frac{0 \cdot 2n^2d^2}{3 \cdot 5d + 8l}\,\mu\text{H.},$$

n being the total number of turns.

This formula obviously takes no account of small variations

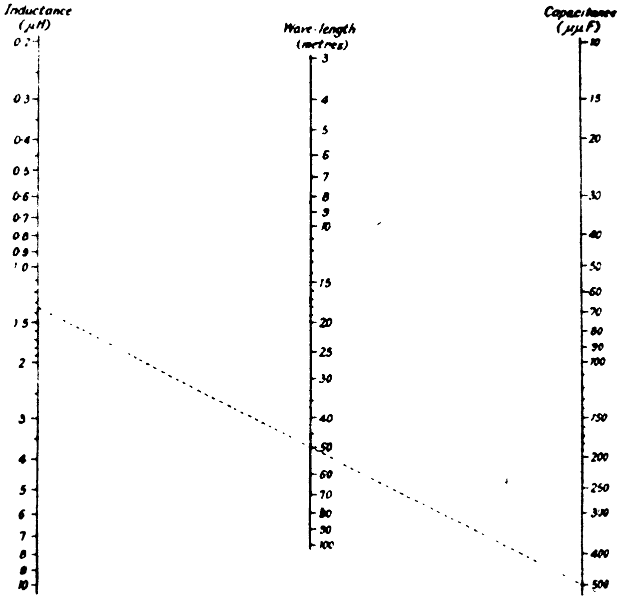

FIG. 65. ABAC FOR DESIGNING SHORT-WAVE COILS

due to thickness of wire, etc., as indicated in Fig. 62, but it is quite accurate enough for most design purposes.

The relation between inductance and the capacitance to tune to a given wavelength may be read off the chart in Fig. 65. By drawing a line between any two of the quantities, the intersection with the third scale will give the quantity required, e.g. to tune to 50 metres with a 0·0005 μF. condenser requires an inductance of 1·35 μH.

Material for Short-wave Circuits.

The valve holders are usually specially constructed of low loss material and in skeleton form, so that there is a minimum of solid dielectric between the various sockets. This can result in quite a marked improvement in efficiency, and a change-over to a valve holder of this type will sometimes permit successful oscillation in a circuit which previously failed to function. In cases where extremely low loss is required, the valve itself is capped in low loss material or is even used without any cap at all. These precautions, however, are usually only necessary on ultra-short waves.

Dielectric Loss at Short Waves.

The question of dielectric loss, i.e. the loss in the former of a coil, the insulation of a condenser, etc., is one which is often imperfectly understood, and it is desirable therefore to examine the problem in a little greater detail.

In a radio receiver we have two types of capacitance. One is the condenser proper, an assembly of parallel plates either fixed or in variable relation, specifically designed to produce a certain capacitance. The second is the capacitance existing between component parts of the circuit. The coils, for example, contain turns of wire at varying potentials and there is a capacitance between such turns, not only between adjacent wires, but between more remote ones, including the terminals at the ends of the winding, the total effect being termed the *self-capacitance* of the coil. In this class we also have the stray capacitances due to wiring, capacitance between the sockets of the valve holders, across the pins of the valve cap, etc. All these unwanted capacitances will pass current to an extent depending upon their position in the circuit and will therefore cause loss.

Now the loss in a condenser is mainly due to a phenomenon known as *dielectric absorption* or *dielectric hysteresis*. It arises from a kind of intermolecular friction which absorbs energy. There is some loss due to the h.f. resistance of the condenser plates themselves and due to any leakage across the condenser due to imperfect insulation, but these effects are usually small.

Experimental evidence shows that the loss due to dielectric hysteresis is inversely proportional to the frequency, which

means that a normal condenser (without undue series resistance or leakage) exhibits a constant power factor.* The power factor is, very nearly, $RC\omega$, which is obviously constant if R is inversely proportional to ω.

Relative Importance of Stray Capacitance.

This means that dielectric loss is of less importance at high frequencies, which would seem to simplify short-wave technique considerably. As against this, we must set the fact that the tuning capacitances in a short-wave circuit are usually much smaller than normal, so that the stray capacitances account for a larger proportion of the total and, since these strays usually have a poor dielectric, the loss is noticeably increased.

It is mainly for this reason that dielectric loss is of importance on short waves. A normal tuning condenser is mostly air-insulated, and remains so until we approach the minimum position, when the inevitable solid supporting material begins to assume an increasing proportion of the total and the loss rises quite sharply. Stray capacitances have the same effect as this solid dielectric in the condenser, and hence should be kept small.

The condenser itself is made with as little solid dielectric as possible, and what material is used is of a high grade. Coil formers may be of special low-loss construction, but as a rule the removal of the former only improves the Q by some 10 per cent, so that special precautions are unnecessary. Valve holders and valve caps, however, may account for 10 to 20 $\mu\mu$F. which is a large proportion of the total capacitance, so that low-loss construction in this direction is helpful.

It is unnecessary to dwell on the subject any longer. If the underlying principles are understood, the relative importance of solid dielectrics in various parts of the circuit will

* Power factor is, strictly, the relation between the true and apparent power in a circuit. The apparent power is the product of the voltage and the current, but this is not the true power unless the two are of similar wave-form and are in phase.

When dealing with two sine waves, the power factor resolves itself into the ratio of the resistance to the impedance. Where the reactive component is large, the ratio may be taken as that of the resistance to the reactance, which for a condenser is simply $RC\omega$.

be appreciated and suitable decisions arrived at for particular cases.

It must be remembered that in transmitting sets the question of losses is more important, for the losses generate heat. Since the power loss in a dielectric rises very sharply with temperature, an excessive loss may produce a cumulative effect, the heat generated warming up the dielectric and causing increase in loss which generates further heat, and so on until the material disrupts.

The table herewith gives an idea of the order of power factor obtained, with different materials at normal temperatures.

Material	Power Factor
Air	0
Mica	0·0004
Ebonite	0·007 to 0·014
Glass	0·004 to 0·016
Bakelite	0·037 to 0·075
Celluloid	0·042
Fibre	0·04 to 0·06
Wood (oak)	0·039
Steatite	0·0005 to 0·001
Rutile	0·0004 to 0·0007

High Frequency Insulating Materials.

A variety of special materials has been evolved in recent years specially for high-frequency work. Apart from rubber-mica mixtures, variously known as loaded ebonite, Keramot or Silvonite, and special plastic materials, similar to bakelite but having lower losses, which are marketed under various trade names, the most important introduction has been the range of ceramics.

These are developed from the well-known electrical porcelain, by suitably changing the constituents to meet the special requirements. One such material is *steatite*, a magnesium silicate mixed with china clay. It is worked in the moist state, as with porcelain, and then "fired."

A second material is *rutile*, a titanium dioxide which has a high dielectric constant. It is thus very suitable for the

manufacture of condensers which are made by depositing a metallic film (usually silver) on to wafers or discs of ceramic insulation. Variable (trimmer) condensers of this type are in large use.

The two requirements are basically different. Steatite and similar materials are employed where insulation is the main requirement. Here a high permittivity (dielectric constant) is not needed and is, in fact, undesirable.

Where the material is required for condenser manufacture, however, the higher the permittivity the less material required for a given capacitance, and the rutile derivatives are particularly suitable. Their main disadvantage is a large temperature coefficient so that the capacitance decreases with increasing temperature, but as the temperature coefficient of other parts of the circuit is usually positive, this effect can be put to good use.

Further information on this subject will be found in papers by W. Jackson, *Journal I.E.E.*, Vol. 79, p. 573, and W. G. Robinson, *Journal I.E.E.*, Vol. 87, p. 570.

Amplification on Short Waves.

For some considerable time it was technically impossible to produce satisfactory radio-frequency amplification at the very high frequencies involved with short-wave working. One of the difficulties is that the smallest stray coupling between anode and grid of the valve will be sufficient to feed back enough energy to cause self-oscillation. Specially neutralized circuits were devised, however, and amplifications of the order of 3 to 5 per stage were successfully achieved. The practice of using neutralized triodes is still in force in transmitting apparatus, as explained in Chapter VI, but for receiver technique the screen-grid valve is almost invariably employed.

With the very small capacitances which are possible with a modern screened tetrode or pentode valve, it is quite practicable to obtain an amplification of 30 to 50 per stage, with medium short waves down to 20 metres or so. Below this, the input impedance of the valves becomes increasingly troublesome, as explained in Chapter IX. The great difficulty, however, is that the amplification is very dependent upon the *LC* ratio of the tuning circuit. Consider a typical case with

an inductance of 2 μH. and a tuning condenser varying from 50 to 250 $\mu\mu$F. The high-frequency resistance would be of the order of 1 ohm, varying over the scale from, say, 1·2 ohms at the high frequency end to 0·5 ohm at the other end. Hence the dynamic resistance (L/CR) would vary from 33 000 to 16 000.

A valve with a slope of 2 would thus give a theoretical gain of 66 at the high frequency end, falling to 32 at the top end. Actually circuit losses would modify this appreciably, particularly at the high frequency end where dielectric losses

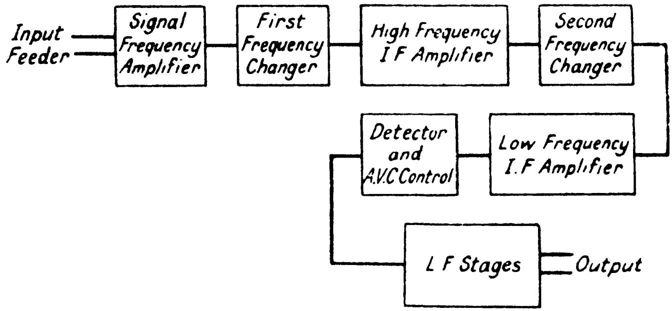

FIG. 66. SCHEMATIC ARRANGEMENT OF COMMERCIAL RECEIVER

would become prominent, and it will be seen that an increase of as little as 1 ohm will reduce the gain to one-half.

Straight radio-frequency amplification, therefore, is little used at short waves. It is often employed as a buffer to separate the aerial from the oscillator of a superheterodyne, this practice having the further advantage that it increases the signal applied to the frequency changer and tends to give a better signal to noise ratio, but the majority of the amplification is usually obtained at some suitable intermediate frequency by using the superheterodyne principle.

Superheterodyne Receivers.

The superheterodyne receiver used for short-wave reception follows out general practice in the main, although advantage is taken of certain factors which enable improved reception to be obtained. The first of these is that a fairly high intermediate frequency is usually employed, which enables

the image frequency to be considerably displaced. With any given setting of the local oscillator, there are two signal frequencies, one on either side, both of which will produce the required intermediate frequency. For example, with a 450 kc/s. i.f., and a local oscillator of 30 Mc/s. we should obtain the required i.f. with signal frequencies of 30·45 and 29·55 Mc/s. which are too close to be separated by a single tuned circuit.

It is possible, by using an h.f. stage in front of the frequency changer, and thereby obtaining two signal-frequency tuning circuits, to reduce the unwanted frequency and strengthen the required channel, but even with such an arrangement the elimination is by no means complete, and the remedy lies either in the use of still more signal-frequency circuits or the choice of a higher intermediate frequency. The latter is the simpler, but suffers from the disadvantage that the gain and selectivity in the i.f. stages both suffer as the frequency is raised.

The broadcast receiver, with its limitations of price and simplicity, compromises by using the 400–500 kc/s. intermediate frequency, which gives fair second channel elimination on short waves, and fair selectivity on medium and long waves. If desired, extra i.f. stages or tuned circuits are incorporated to raise the standard of selectivity.

For a commercial receiver, greater elaboration is possible. It is practicable, for example, to convert first to a high intermediate frequency of perhaps 2 000 or 3 000 kc/s. and incorporate two or three stages at this frequency. The selectivity and stage gain will not be great, but they will serve to eliminate practically entirely any second channel (image) interference. A further frequency changing stage is then included which feeds an amplifier operating at a much lower intermediate frequency where the selectivity and gain are much greater. This amplifier would probably contain band filters, i.e. filters which accept all frequencies within a certain range, but cut off on either side of this range very sharply. Such filters are arranged by a combination of high- and low-pass filter networks such as are discussed in *Modern Radio Communication*, Vol. II, and while quite impossible on the score of cost in a broadcast receiver, are more than justified in commercial practice.

The output from the last i.f. valve is applied to a suitable detector which feeds the audio-frequency stages. These may take several forms. For speech reception a standard low-frequency amplifier is employed. For telegraphy reception the signals are suitably amplified and are then applied to further rectifier valves, which convert the audio-frequency signals into a d.c. pulse, which is passed on to a recorder.

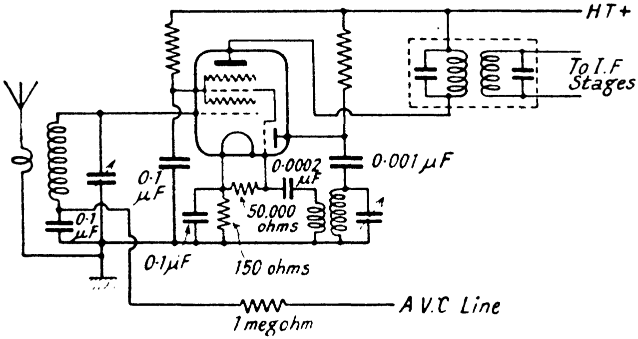

FIG. 67. TRIODE-HEXODE FREQUENCY CHANGER

There are various ways of receiving this, which need not be elaborated in detail, the essential principle being the same throughout.

It is often useful in long-wave practice to tune the audio-frequency amplifier in such cases so that it gives a maximum response to audio-frequency signals of particular note of the transmission being received. This provides a further measure of selectivity by excluding unwanted transmissions and atmospheric disturbances to some extent, but such measures are not usually necessary on short waves where adequate selectivity can be obtained by radio-frequency methods.

The Frequency Changer.

The circuits used for converting the signal frequency to intermediate frequency are similar to those employed at medium and long wavelengths, with certain reservations. The simple electronically-coupled type of valve is unsuitable, because the capacitance between the electrodes inside the

valve provides a coupling between the aerial and oscillator circuits. While this is very small in terms of capacitance, it is quite large enough at the high frequencies involved to cause "pulling" between the two circuits. They form a tight-coupled arrangement, and any variation of the aerial circuit affects the oscillator and vice versa. Since, in order to obtain the necessary frequency-changing action, the oscillator must differ from the signal frequency by the required amount, this means that under operating conditions the signal frequency circuit is always mis-tuned.

Short-wave practice, therefore, prefers a valve of the hexode type, in which the mixing grid is in the electron stream and capable of modulating it as before, but is screened both from the anode and from the signal grid as indicated in Fig. 67. With such a valve, if the circuits are suitably screened from one another, the interaction is negligible.

The hexode may be used in conjunction with an entirely separate oscillator or in the form of a combined valve known as a *triode-hexode*.

Frequency Drift.

Trouble arises at short waves due to variation in the frequency of the local oscillator. Part of this is caused by changes in the inductance and capacitance of the oscillator circuit, but by suitable design, including the use of temperature-compensated condensers, this may be largely eliminated. The remainder arises from a change in the effective capacitance of the valve. This is two-fold. Part arises from changes in the valve itself which can be reduced by suitable design. In particular, the heater wattage of the modern valve has been cut down by more than 50 per cent. The second effect is due to alteration of operating conditions by the application of a.v.c. (automatic volume control).

The electrons which pass the first screening or accelerating grid (G_2) enter a retarding field due to the injector grid (G_3), which is at a negative potential. Hence a space charge accumulates which increases the effective capacitance between G_3 and earth. This space charge, however, varies with the bias on the control grid G_1, so that the application of a.v.c. to this grid causes a change in oscillator frequency. (See Fig. 68.)

If G_3 is of wide mesh, requiring a large oscillator voltage, the effect is small, but since the generation of large oscillating voltages is difficult at short waves, attempts have been made to improve matters by using a closer mesh. This has increased the effect of G_3 on the electron stream but has also increased the space charge and the frequency variation in consequence. The use of a tuned anode oscillator circuit minimizes this effect, but the difficulty still remains.

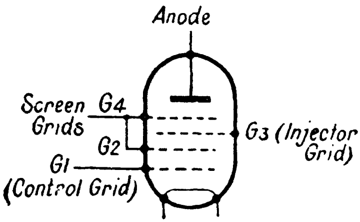

FIG. 68

With commercial technique, when cost is of secondary importance, the trouble can largely be obviated by using a separate oscillator valve of higher slope with a hexode having an open-mesh injector grid, and using little or no a.v.c. on this valve; but for broadcast receivers this solution is impracticable and attention has been refocused on the octode type of mixer.

Octode Frequency Changers.

Such a valve is illustrated in Fig. 69. The first and second grids are used as the triode for maintaining the oscillation,

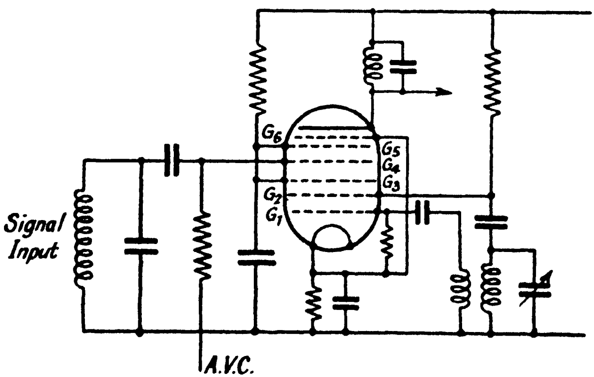

FIG. 69. OCTODE MIXING CIRCUIT

while the capacitance between the signal grid and the oscillator

is reduced by interposing another screen or accelerator grid G_3 between G_2 and G_4.

This, however, is only partially successful. For a stable oscillation the oscillator anode current must be in phase with the grid voltage. If it lags behind its correct phase, the oscillation frequency will be less than the resonant frequency of the

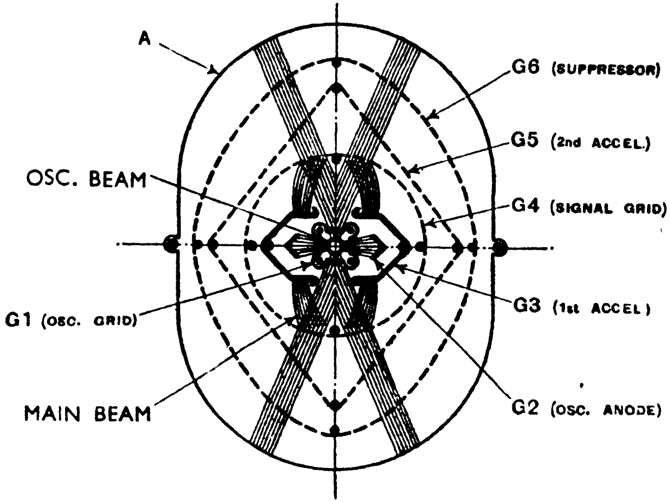

FIG. 70. ARRANGEMENT OF ELECTRODES IN "BEAM" OCTODE

tuned circuit by an amount necessary to bring the phases correct. But under such conditions the frequency is dependent on the phase lag, which is controlled by the valve, so that the system is prone to drift. Phase lag occurs to an increasing extent at high frequencies owing to the finite transit time of the electrons, and, as many of these electrons reach the oscillator anode by devious routes, the effect is noticeable as low as 10 megacycles.

This difficulty has been overcome by the application of electron optics to valve construction, the electrodes being so disposed as to direct the electrons in beams where they are required. Fig. 70 illustrates a typical arrangement. G_1 is provided with four solid pillars, which only allow the electrons to

escape in four beams. Two of these are collected almost entirely by the oscillator anode, and the transit time of this section is thus very short.

Around this is the first accelerator G_3, which is solid and effectively screens the oscillator, except for two small gaps which permit the other two beams to pass. These beams, which are modulated by the oscillator grid G_1, then pass through G_4, which again modulates them and produces the required mixing, and then on to the main anode. Such of these electrons as are returned because of the negative potential of G_4 are collected by G_3 and thus prevented from interacting with the oscillator.

Short-wave Valves.

Apart from these matters, however, it will be clear that there are other respects in which the design of the valve requires special attention for short-wave operation. One of the major difficulties is that of transit time, which is more fully discussed in the next chapter.

In brief, at high frequencies the time taken by the electrons to move between the electrodes of the valve is no longer negligible, so that simple explanations of the action of the valve no longer apply.

In the case of the hexode frequency changer, for example, if the time of one oscillation is comparable with the time taken by the electrons to travel from G_1 to G_2 (Fig. 68), the electrons will be driven back to the control grid G_1. This will build up a charge on this grid and (since it has vari-mu characteristics) will reduce the amplification. Reduction in the clearance between G_1 and G_2 minimizes this effect so that with a modern valve the gain can be maintained up to 30 Mc/s., but beyond this the falling off is still apparent.

Another difficulty arises from the length of the leads from the valve, particularly the cathode lead. This lead has inductance which is common to both anode and grid circuits. Hence the a.c. anode current will develop voltage across this inductance which is transferred through the cathode-grid capacitance to the grid, and these voltages are in such a direction as to reduce the effective amplification and increase the apparent input resistance of the valve.

In fact, as the frequency increases, this effect becomes of

more importance than the Miller effect discussed on page 87, and a low grid-cathode capacitance is of more value than a low anode-grid capacitance.

To meet these requirements, special "ring seal" methods of manufacture are being adopted, and will probably become universally used for receiving valves. This construction reduces the length of lead to a small fraction of an inch and also materially reduces the grid-cathode capacitance.

Tracking.

The signal frequency and oscillator circuits are usually ganged together, the oscillator circuit being padded so that the frequency at any given setting of the oscillator condenser is just the required amount higher than the tune of the signal frequency circuit at the same setting of the condenser. This may be accomplished by a circuit of the type shown in Fig. 71.

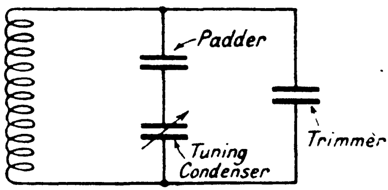

FIG. 71. TRACKING CIRCUIT FOR SUPERHETERODYNE OSCILLATOR

If we use the higher of the two possible oscillator frequencies, then the frequency range from the oscillator is slightly less than that of the signal frequency circuit. The capacitance range of the condenser should therefore be slightly reduced, which is done by including a fixed *padder* condenser in series with the main tuning condenser. Then if the inductance of the coil is suitably chosen, we can arrange the frequency coverage to be just what is required.

An alternative method of achieving the same result is to restrict the capacitance range by adding a parallel *trimmer* condenser and again choosing the inductance to suit. The inductance in this second case will be different from that in the first.

Both these methods will give correct operation at the two ends of the scale, but it does not follow that it will be correct in between, and actually there will be an error which we find to be positive with one system and negative with the other.

We can, therefore, by using both methods, obtain an arrangement which gives nearly correct tracking over the whole scale. For medium and long waves the calculations involved are somewhat elaborate, but for short-wave working it is usually sufficient to use the trimming condenser only, the error introduced by dispensing with the padding condenser being usually small enough to be neglected. This desirable state of affairs arises from the fact that the frequency difference is usually quite a small percentage of the frequency itself, under which conditions the padding condenser comes out at a very large value, and the error introduced by making it infinitely large, i.e. short-circuiting it, is negligible.

On this basis the calculation can be simplified as follows—

Let L_s and L_o be the signal frequency and oscillator inductances respectively;

Let C be the maximum capacitance of the tuning condenser (including stray and circuit capacitance).

Then assuming the same stray and trimmer capacitances in both signal and oscillator circuits, we have

$$f_s = 1/2\pi\sqrt{(L_s C)}$$
$$f_s + f = 1/2\pi\sqrt{(L_o C)},$$

f being the intermediate frequency.

Whence $$L_o = L_s f_s^2/(f_s + f)^2.$$

Actually the trimmers are not the same for both circuits, but only a small error is involved by assuming that they are.

We now can find the capacitance needed to tune L_s to the maximum frequency (minimum wavelength) required. This will be obtained from the minimum of the condenser C_m plus a suitable trimmer capacitance C_1 (which will include strays). The oscillator capacitance is then easily shown to be

$$C_o = \frac{f_s^2 L_s}{(f_s + f)^2 L_o} \cdot C_s.$$

This capacitance C_o will be made up of the same condenser minimum C_m plus a trimmer. Hence the oscillator trimmer $= C_o - C_m$.

In practice, this second stage of the calculation can often be neglected. Experience shows that the oscillator trimmer

will have to be a little larger than that across the signal-frequency circuit, and allowance is made accordingly, so that it is only necessary to calculate the oscillator inductance.

Super-regenerative Receivers.

A type of receiver which has some possibilities for simple reception on short waves is the super-regenerative type. This form of receiver includes a simple regenerative or reacting circuit having a very smooth reaction control, so that it slides into oscillation extremely easily. Now actually when a receiver starts to oscillate it does not immediately attain a steady condition. Let us suppose that the reaction setting is such as to permit the circuit to oscillate, but that there is no oscillation present. Some momentary disturbance upsets the equilibrium and a minute oscillation starts. Due to the reaction, this oscillation builds up amplitude, and will continue to grow until the oscillating current is such that the losses are just equal to the energy supplied by the valve.

This steady state oscillation will be considerably greater than the current at the threshold of oscillation, which is the maximum current normally available in a reacting receiver. The process of super-regeneration involves the quenching or stopping of the oscillation before it has time to reach its full value of oscillation. Under these conditions the amplitude which it will attain during the period over which it is allowed to grow will depend upon the strength of the initial oscillation, so that we attain an amplitude proportional to the incoming signal, but much greater than is possible with the simple reacting arrangement.

The quenching, of course, continues indefinitely, so that the oscillation is repeatedly checked and then allowed to build up again; and if the circuit has been correctly designed, the final amplitude of the oscillation will all the time follow the modulation imposed by the incoming signal. It will be clear that this involves a quenching frequency higher than the highest modulation frequency. Unfortunately, this does not always allow the signal adequate time to build up, and on long waves a high audio frequency has to be used, giving a continuous "sing" which is objectionable. On short waves,

however, it is possible to quench at a frequency well above the audible limit.

The circuit is capable of very high amplification, particularly with extremely weak inputs. It is difficult to adjust, for it depends essentially upon strict proportionality between the oscillation level actually attained and the modulation of the incoming signal, and it has the defect of being very unselective.

Receiving Wavemeters.

For estimating the frequency of receiving circuits the requirements are less stringent than with transmitters. An absorption wavemeter of the type described on page 107 is usually used, and the frequency of an oscillator may be determined by including a milliammeter in the anode circuit of the oscillator valve. If the circuit ceases to oscillate a change in the anode current will be observed, and by bringing the wavemeter near the oscillator coil a flicker of the meter needle will be observed when the circuits come in tune. The wavemeter should be coupled as loosely as possible so that it only just affects the oscillator. Otherwise the calibration of the wavemeter may be altered.

For calibrating non-oscillating circuits, voltage is injected from a suitable calibrated source and the voltage developed is observed with a suitable indicator which may be a valve voltmeter or some part of the set itself. For example, if the input is modulated, an audio-frequency voltage will appear at the output of the set which may be detected aurally or by meter, and this is a maximum when the circuits are in tune.

The local source may be a signal generator, i.e. a calibrated screened local source, usually modulated to a known degree and provided with a calibrated attenuator for controlling the output, or for more accurate work, a heterodyne wavemeter consisting of a carefully designed circuit maintained in oscillation by means of a valve so coupled as to have negligible influence on the calibration.

CHAPTER IX

ULTRA-SHORT WAVES

WE have already seen in Chapter II that the ordinary mechanism whereby the wireless waves are reflected in the upper atmosphere breaks down when the wavelength approaches 10 metres. At this point the bending of the ray is barely sufficient to enable it to return to Earth, and a point is reached at which the waves fail to return altogether, so that they are lost. The critical wavelength is in the neighbourhood of 10 metres, although, in the case of most phenomena connected with the upper atmosphere, no sharp division is obtained. Under certain conditions, wavelengths of less than 10 metres will be successfully reflected, while at other times wavelengths of greater length, even up to 12 metres, will not get through. Ordinary short-wave transmission therefore does not use wavelengths lower than about 13 metres, to allow a reasonable margin of safety.

Once again, therefore, we reach an apparent barrier to further progress, but, as in other parts of the spectrum, the barrier is neither sharp nor complete. Duly authenticated evidence exists of the reception of ultra-short waves over thousands of miles. Such reception is erratic, however, and it is generally accepted that waves below 10 metres can only achieve, with reliability, a more or less optical range.

Ground-ray Transmission.

Let us consider the ground wave first. Just as one can see farther from a high point in the landscape, so the range of an ultra-short wave transmitter increases with the height of the transmitting aerial. The optical range (i.e. the distance from the transmitting point to the horizon) is easily shown to be

$$d_{(miles)} = 1{\cdot}22\sqrt{(\text{height in feet})}.$$

The curve of Fig. 72 shows this range plotted in terms of the height, and illustrates the rapid rise at first, followed by

a more gradual increase afterwards which one would expect from the formula just quoted.

Now the radiation from an aerial in free space produces a field strength which falls off inversely as the distance. The farther away we go the less is the strength, in strict proportion. With a simple half-wave aerial, the field is given very simply by $E = 7\sqrt{P}/d$, where P is the radiated power in watts.

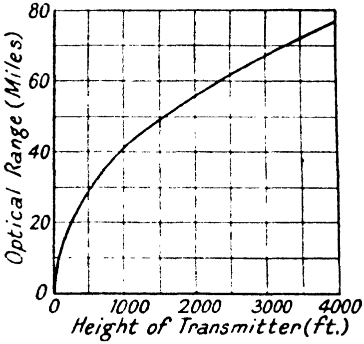

FIG. 72. VARIATION OF OPTICAL RANGE WITH HEIGHT

With a practical transmission, however, the wave at the receiving point is complex, and arrives by two main paths. The first of these is the direct route from the transmitter. The second is a wave which reaches the ground just a little in front of the receiving point, and is reflected off the ground. This wave strikes the ground and is subsequently reflected at a very small angle—in fact, it just grazes the ground—and thus is still at ground level and will affect the receiving aerial together with the direct wave. (Waves striking the ground much earlier will, of course, do so at a sharper angle, and the reflected wave will pass over the top of the receiving aerial.)

Now this grazing wave is reflected with a high efficiency, but, as with any reflection, suffers a phase reversal, so that it is in opposition to the direct wave and thus reduces the effective field strength. For flat ground, Trevor and Carter (*Proc. I.R.E.*, March, 1933) have shown that the field strength becomes

$$E = 88h_1h_2\sqrt{P}/\lambda d^2 \text{ volts per metre,}$$

where h_1 is the height of the transmitting aerial

h_2 is the height of the receiving aerial

λ is the wavelength

d is the distance, all in metres

and P is the radiated power in watts.

It will be seen that we now have an inverse square law, so that if we double the distance the field strength will be reduced to one-quarter. In practice, the ground is not flat, but may undulate, while buildings and similar structures will modify the reflections so that the actual signal strength is less than the theoretical amount, but practical experience tends to confirm the inverse square law, the observed field strength never exceeding the theoretical value (except in isolated cases where the receiver is considerably higher than the surrounding country), and being generally between the theoretical value and one-tenth of that value with an average of about one-third.

Propagation Beyond the Horizon.

The signal does not cut off sharply, however, beyond the horizon, but is still evident, though at rapidly decreasing strength. This is due to *diffraction*. The dragging of the feet of a wireless wave due to the resistance of the earth has already been referred to (page 71), and this effect now comes into play to keep the wave, or a portion thereof, on the earth instead of leaving it altogether.

The strength of the wave now falls off much more rapidly, but signals can still be detected at ranges two or three times the optical distance. Beverage quotes a series of measurements in the *R.C.A. Review*, January, 1937, which show, firstly, a normal inverse-square law relation up to the horizon and then a sharp discontinuity, followed by an attenuation at a much more rapid rate depending on the wavelength. With waves around 10 metres, the post-horizon attenuation is roughly proportional to $d^{3\cdot 2}$, while at wavelengths of the order of 1 metre the attenuation is proportional to d^8—a very marked increase.

Since only four sets of measurements were involved, Beverage quotes the results with reserve, but it is clear that for transmission over the greatest "ground" distance the longer waves are preferable.

Refraction.

There is also evidence of a refraction or bending of the waves in the atmosphere, so that waves which would normally miss the Earth altogether may be bent round and reach the

ground at some point beyond the horizon. Such waves will interfere with the diffraction waves just discussed and produce fading.

Ross Hull (*Q.S.T.*, June, 1935) has shown that this reception is usually obtained under conditions of *temperature inversion*. Normally, the temperature of the air falls roughly 1° C. for every 300 ft. above the Earth's surface, and this fall continues more or less uniformly throughout the troposphere. Under certain atmospheric conditions, however, the temperature in some localities will rise with increasing height. The rise is limited and is ultimately, of course, succeeded by a rapid fall to restore the equilibrium condition, but if such an area happens to be located in between the transmitter and receiver, refraction occurs.

Long-distance Reception.

Finally, there are instances of ultra-short wave reception over thousands of miles. Such reception is erratic and little can be said about it until more data are available. It can only be explained by some form of reflection, and this has hitherto been assumed to be due to abnormal conditions of high ionization in the E or F layers.

As explained in Chapter II, Watson Watt in 1937 postulated the existence of three further electrified layers *below* the E and F layers.

These layers occur in the troposphere and are atmospheric rather than electronic in character. Their existence appears to be well established and information is steadily being compiled as to their character and reliability. It does not appear likely that long distance transmission with ultra-short waves will prove reliable, and the principal effect of these low-level layers is to introduce variable factors into the relatively uniform conditions appertaining to the ionospheric reflections utilized in the transmissions with longer wavelengths.

Advantages of Ultra-short Waves.

The limitations of ultra-short wave transmissions might appear to render them of little value. Such, however, is not the case. Even the restriction of the range to some small multiple of the optical path has distinct advantages. Apart

from the absence of fading (except in the post-horizon region), there is the important feature that interference from other stations is eliminated, provided that such stations are outside the range of reception. This is an advantage, because it means that a number of stations separated by some hundreds of miles can all operate on the same wavelength without interference between one another, and in view of the increasing demands made upon the available space in the ether, this is likely to prove of considerable advantage. A further advantage which we shall discuss more fully later in the chapter is that, owing to the very high frequency of the carrier employed, it is possible to modulate it satisfactorily with frequencies as high as 2 Mc/s., a fact which is of considerable assistance in television transmission.

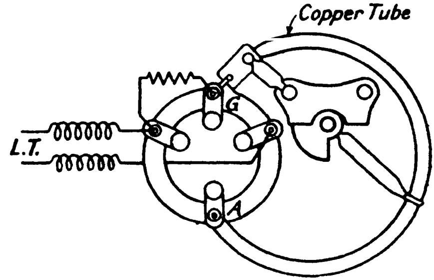

Fig. 73. Lay-out of Simple U.S.W. Oscillator

U.S.W. Transmitters.

Transmitters for ultra-short waves can adopt ordinary technique down to somewhere between 2 and 3 metres. The circuit adopted has to be very symmetrical, and the length of the leads constitutes a considerable proportion of the circuit. The most successful arrangement, therefore, is to make the leads themselves form the circuit. Fig. 73 shows the lay-out of a 5-metre oscillator constructed on this idea. High-frequency chokes are often included in the filament leads and in some of the supply leads in order to prevent leakage of the radio-frequency energy in unauthorized channels. These chokes consist of five to ten turns of wire on about a 1 in. diameter former having an inductance of a few microhenries only. This is sufficient to present a very high impedance to the frequencies being generated, and therefore acts as an almost complete barrier.

The usual difficulties with frequency stability are encountered. It is not desirable, except in the simplest circuits,

to couple the aerial direct to the transmitter, since this limits the power which can be extracted from the transmitter without instability (see page 81) and, secondly, any variations in the aerial system will be communicated to the transmitter and thereby cause the frequency to drift. As against this, ultra-short wave aerials can be made very rigid, and by using copper tube fixed at both ends and in the middle to a suitable framework, aerial variations can be greatly minimized, and for small transmitters the transmitter itself can actually be built at the centre of the aerial system, thereby avoiding the necessity for any feeders.

Fig. 74 shows a push-pull circuit which can be successfully used for a transmitter of this character. The advantage of the push-pull is that still greater symmetry is obtained and, of course, more power.

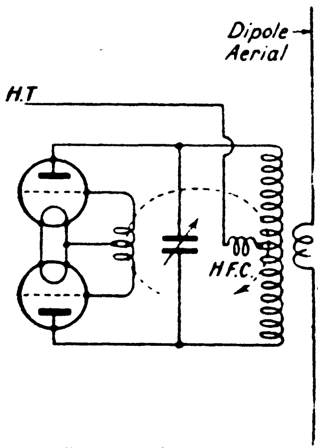

Fig. 74. Push-pull Transmitter

For more elaborate transmitters, however, amplifiers must be used. These amplifiers will still give a small gain, although at wavelengths as low as 2 to 3 metres the amplification may fall practically to unity, so that the valve acts merely as a buffer rather than an amplifier. If the transmitter is located at any distance from the aerial, the two must be connected by suitably designed feeders in accordance with the data already given in Chapter III. The greatest care, however, must be taken to avoid unwanted radiation from the feeder. Some authorities maintain that the feeders must be completely screened for satisfactory results.

Wavelength Measurement.

For measuring the wavelength or frequency of ultra-short waves, methods may be adopted similar to those employed for longer waves. The increasingly short wavelengths involved,

however, make it possible to determine the wavelength by actual measurement, using a tuned feeder or Lecher wire, as explained on page 182. Both methods are in use, according to requirements, a wavemeter being more convenient under certain conditions and a Lecher wire in other circumstances.

Frequency Stability.

It is important that the frequency shall not drift or wander. Stabilization of the order obtainable with normal short waves

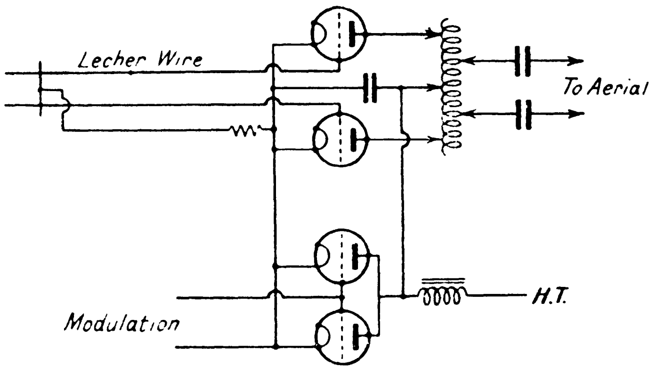

FIG. 75. PUSH-PULL TRANSMITTER WITH LONG-LINE STABILIZATION

is not possible, for crystal control or similar methods are impracticable owing to the high frequencies involved. Vibration in the aerial due to wind can cause appreciable variation, and the structure must therefore be as rigid as possible.

A master oscillator feeding an amplifier is preferable when conditions permit, even though the gain is probably little more than unity, and, of course, the component parts of the transmitting circuit must be constructed on lines which will minimize variations due to temperature, etc.

A method which has been used successfully is that known as *long-line stabilization*, illustrated in Fig. 75. Here a push-pull oscillator is employed with a tuned grid circuit in the form of a Lecher wire. Any variation in frequency causes rapid change in the amplitude and phase of the grid voltage due to modification of the standing wave on the wires, and a

stability of a small fraction of 0·1 per cent is obtainable within normal variations of temperature. The Lecher wire, in fact, acts as a high Q circuit, giving the optimum conditions for frequency stability, as explained in Chapter VI.

Limit of Valve Operation.

It is found increasingly difficult to produce oscillation with a valve of normal construction as the frequency is raised, while any attempt to obtain radio-frequency amplification meets with failure. Even when all precautions have been taken to avoid stray capacitances, to neutralize any feedback and so forth, the effective amplification is found to be negligible. In fact, in many cases the stage steps down rather than up.

The reason for this is to be found in the fact that the time of transit of electrons across the inter-electrode space is comparable with the period of the oscillation being amplified or generated. At first this might appear to result merely in a small delay in the necessary operation, but unfortunately the effect is rather more troublesome than this. Under normal conditions we have a stream of electrons flowing from cathode to anode. The grid is biased negatively to such an extent that it absorbs none of these electrons—i.e. no grid current flows, and the electron stream is controlled by the variations in the potential of the grid above or below this mean value. At ordinary frequencies this variation is slow compared with the movement of the electrons, and a satisfactory and smooth control is obtained.

When the frequency is very high, however, a different state of affairs results. Suppose, for example, that we suddenly make the grid much more negative. The increased grid potential will prevent any further electrons from being emitted from the cathode, but we are still left with a quantity of electrons which have already left the cathode, and are prevented from reaching the grid due to the fact that it is now at a negative potential. Similarly, the electrons in between the grid and the anode are suddenly repelled, and we get a strained and unnatural condition which is immediately neutralized by a flow of electrons through the external grid circuit, the direction of this current being the same as if the grid-cathode space were conducting.

A similar but opposite state of affairs would result if the grid were made suddenly positive, so that we obtain in actuality a momentary grid current, despite the fact that the grid is so biased that under static conditions no grid current flows. If the grid is being subjected to extremely rapid alternations of potential, a permanent alternating grid current will flow exactly as if there were a resistance across the grid and cathode. In other words, the input impedance of the valve falls from a very high value at normal frequencies down to something quite small—of the order of 10 000 ohms or even less—at frequencies comparable with the transit time of the electrons.

Input Conductance.

The subject was discussed very fully in a paper by Ferris (*Proc. I.R.E.*, Jan., 1936), in which he shows that the input conductance is given by the expression

$$g = Ksf^2\tau^2$$

where s is the slope of the valve,
f is the frequency,
τ is the electron transit time,
and K is a constant depending on the geometry and operating conditions of the valve.

An evaluation of this expression shows that with an ordinary valve the input resistance may be as low as 20 000 ohms even at 30 Mc/s. (10 metres), and the results are confirmed by actual experiment. It will be noted that in a given valve τ is constant, so that the grid conductance increases as the square of the frequency, and thus very rapidly passes from a point of negligible importance to one of major consideration. (It must be remembered that conductance is the reciprocal of the input impedance, so that with a perfect valve we require the conductance to be zero, and the higher the input conductance, the worse the valve.)

Acorn Valves.

One remedy is to reduce the transit time of electrons, which is done by making special valves having extremely small clearances. Such valves are now available and are in commercial

use. They are barely ½ inch in diameter and are not provided with any cap in the ordinary sense of the word, the leads being brought out through the sides of the bulb as indicated in Fig. 76. The general appearance of the valve is rather similar to an acorn, which has resulted in the name *acorn tubes*. The reduction in the size unfortunately tends to increase the slope, which in turn increases the grid conductance, but by correct design it is possible to make the decrease in τ^2 outweigh the increase in slope by a considerable margin, and with such a valve the input resistance can be kept as high as some 50 000 ohms even at frequencies between 50 and 60 Mc/s.

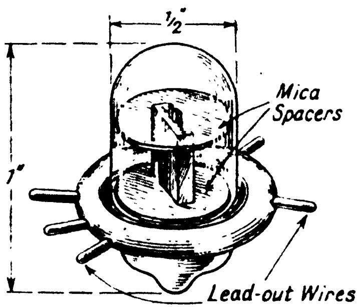

FIG. 76. ACORN VALVE FOR ULTRA-SHORT WAVES

Satisfactory and commercial amplification at wavelengths of 7 metres or less has been obtained using valves of this type.

The effect of the grid loss can be minimized by tapping the grid down the tuned circuit as is done to minimize the detector damping in normal receivers. This is a well-known technique and need not be discussed here further. For more detailed information the reader should refer to *Modern Radio Communication*, Vol. II.

Transmitting Valves.

Similar precautions are necessary with transmitting valves with the added proviso that the output obtainable is limited by the power which the valve itself will dissipate. Even if the structure of the valve can be improved in this respect it is still necessary to bring out leads to the external circuit.

But at the high frequencies involved these leads constitute an appreciable part of the circuit and carry heavy currents. Both grid and anode leads therefore are relatively massive, and are brought direct through the glass envelope at the side or top. The limit of output with a glass valve, however, is about 50 watts at 300 Mc/s.

Better results are obtainable by adopting the long-wave technique of making the anode the outer container, the grid and filament being housed on a glass support fixed to the anode with a glass-to-metal seal as shown in Fig. 77. Cooling may be assisted by fitting radiating fins as shown in the figure, by forced air draught or by water cooling. By the last-mentioned technique up to 150 kW may be dissipated at 25 Mc/s and up to 2·5 kW at 300 Mc/s.

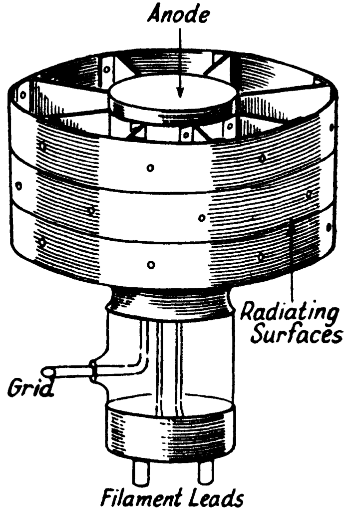

FIG. 77. COOLED-ANODE TRIODE

U.S.W. Receivers.

Receivers for ultra-short wave reception follow customary practice, again with due allowance for the increased frequency. Simple regenerative receivers are only suitable for large signals, while the super-regenerative type discussed in Chapter VIII has met with some favour. As the wave-length is reduced so that the frequency of the carrier increases, it is possible to get a satisfactory super-regenerative action with a quenching frequency well above the audible limit, so that the usual "sing" is eliminated, and owing to the relative freedom of ultra-short waves from atmospheric disturbances, the background noise is not so bad.

The super-regenerative circuit, however, has never been a very well-behaved arrangement, being capable of extremely good performance under proper conditions, but giving very ordinary results if the circuit for any reason is not functioning properly. The majority of receivers employed for this class of reception, therefore, are of the superheterodyne type. The intermediate frequency chosen depends entirely on conditions. If only sound reception is to be obtained so that the modulation frequency does not cover more than 10 or 12 kc/s. on each

side of the carrier, then an intermediate frequency of normal character can be used. Selectivity is not a troublesome problem owing to the extremely limited number of stations operating on ultra-short waves, and the absence of interference from any station more than 100 miles or so distant.

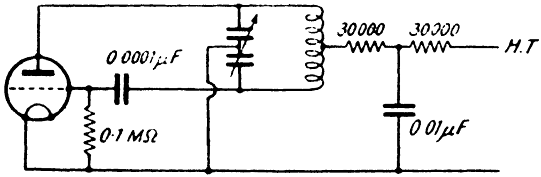

FIG. 78. COLPITTS OSCILLATOR FOR ULTRA-SHORT WAVES

For the reception of television signals it is necessary to use special intermediate frequency amplifiers capable of handling the very wide band width necessary, and this point is discussed in more detail later in the chapter.

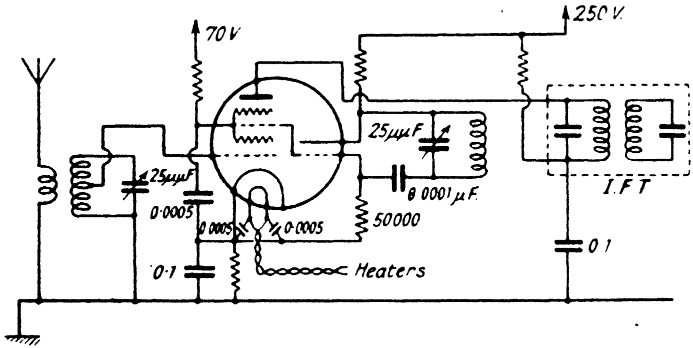

FIG. 79. ULTRAUDION FREQUENCY CHANGER FOR ULTRA-SHORT WAVES

A point of considerable importance on these wavelengths, however, is the frequency changer. While the customary simple method may be used such as the triode-hexode of Fig. 67, the conversion gain with such an arrangement is liable to be somewhat limited. A further difficulty is that parasitic oscillations are often obtained which render the

whole circuit lifeless. This trouble can be avoided by using a Colpitts or ultraudion oscillator as shown in Figs. 78 and 79. In both cases, of course, the valve shown may be the triode section of a triode hexode, and Fig. 79 shows the complete circuit.

Note the bypassing of the heater connections with small (mica) condensers and the tapping of the signal grid down the coil to minimize valve damping, as already explained.

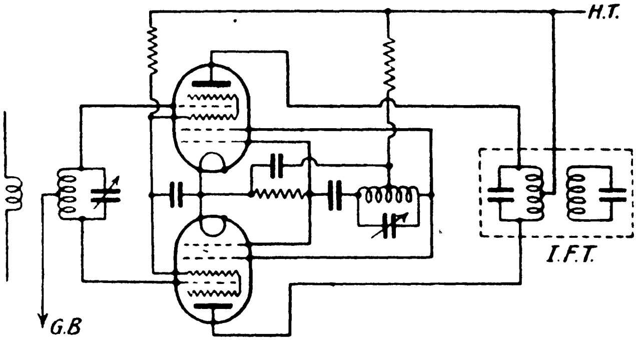

FIG. 80. PUSH-PULL HEPTODE FREQUENCY CHANGER

An alternative arrangement is to use a push-pull circuit as in Fig. 80. Each valve is tapped across half the coil, and though both valves are in operation together, the total damping is reduced and increased gain results, while the symmetry of the circuit is of material advantage.

The coils for ultra-short wave working are, of course, smaller than those for the normal short-wave band, and the greatest care must be taken to keep all leads as short as possible, as otherwise the inductance and self-capacitance of the leads themselves will more than provide the necessary tuning characteristics.

The aerial is usually of the half-wave type connected to the receiver through a suitable feeder, usually of the tuned variety. Ultra-short waves are liable to marked reflection from natural objects or buildings, and the aerial should be located in as open a space as possible. A signal may be almost

unreceivable at ground level, but of quite good strength 20 or 30 ft. up.

Interference.

Atmospheric disturbance is negligible on these very short waves, but serious interference is caused by the ignition systems of motor cars which radiate on a wavelength around 7 metres. The intensity of this interference increases as the receiving aerial is raised above the ground, but after a point it falls off again rapidly, as shown in Fig. 81. Since the signal usually increases with the height of the aerial, the signal-to-noise ratio shows a marked improvement above about 10 ft.; and if, at the same time, the aerial can be located some distance away from any road—20 ft. or so will usually suffice—reasonably clear reception can be obtained.

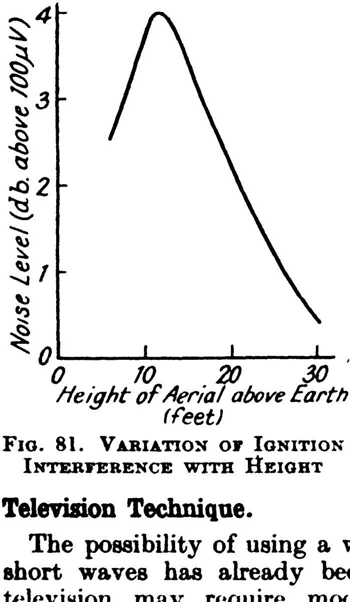

FIG. 81. VARIATION OF IGNITION INTERFERENCE WITH HEIGHT

Television Technique.

The possibility of using a wide modulation band on ultra-short waves has already been referred to. High-definition television may require modulation frequencies exceeding 5 Mc/s., and hence any medium or long-wave transmission is quite out of the question, since the modulation frequency would be greater than the carrier. Even short waves are not suitable, owing principally to the selective fading which has already been mentioned in Chapter VII (page 114).

All the transmissions of to-day, therefore, are taking place on ultra-short waves, usually at frequencies between 40 and 50 Mc/s. With such carrier frequencies a modulation of 2 to 5 Mc/s. is only a small percentage of the carrier, and there is no serious difficulty in transmitting the required intelligence.

The modulating circuits at the transmitter must, of course,

be specially designed to handle such very high frequencies. A series modulating circuit is often employed for the reason that it involves no anode impedances, so that a uniform modulating characteristic is obtained up to the limit where valve capacitances begin to be comparable with the internal resistance of the modulating valve.

Similarly all the r.f. tuned circuits must be designed to have a wide band spread as in the case of the receiving circuits, as discussed in the next section.

Reception.

Television reception requires special technique due to the wide band width required. Even with a 50-Mc/s. carrier the response of a normal circuit will be 50 per cent down 0·5 Mc/s. off tune. Conditions are improved by the low input resistance of the valves at these high frequencies. This flattens the tuning considerably at the expense of the stage gain, but even then the loss of the higher modulation frequencies is quite appreciable.

It is usually necessary to shunt the circuit with a low resistance of the order of 2 000 ohms in order to obtain adequate spread and in these circumstances ordinary valves may be used with better results than special "acorn" types.

The alternative method is to adopt the superheterodyne principle, but it is clear that a special type of i.f. amplifier will be required, for the resonance curve of the customary i.f. transformer only has a band spread of 10 to 20 kc/s. It is customary, therefore, to use an intermediate frequency of 5 to 10 Mc/s. A suitably coupled band-pass circuit will then give an acceptance of the order required. Fig. 82 shows a typical response curve of a television i.f. transformer.

Even with band-pass coupling, resistance damping may be required and in many cases designers prefer to use single circuits shunted by, say, 1 000 ohms which gives nearly as good a characteristic. Whichever system is used, the gain per stage is limited by the wide band width required to about 5.

The precautions required in connection with the frequency changer have already been mentioned. The detector stage can be a diode feeding into video-frequency amplifiers (i.e.

low-frequency amplifiers capable of handling the high frequencies required for vision), and it is a matter of preference as to whether the high-frequency amplification shall be continued until the diode is able to load the cathode ray tube or other receiving device directly, or whether a smaller detector input shall be used with a suitable number of video-frequency stages.

Video-frequency Amplification.

Any amplification subsequent to the detector must again be of a special character capable of transmitting all frequencies

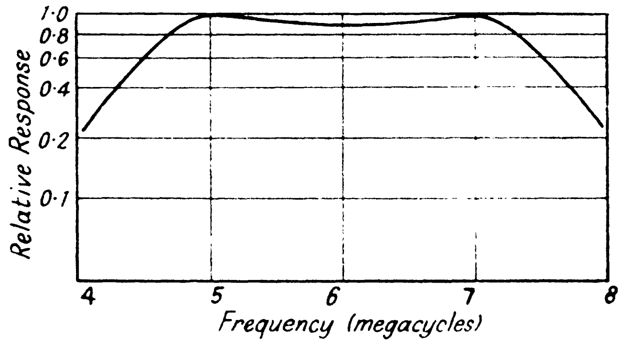

FIG. 82. RESPONSE CURVE OF TYPICAL TELEVISION I.F. TRANSFORMER

from 25 c/s. to 2 Mc/s. or more. This may be accomplished with resistance coupled screened valves, but the anode resistances must be such that they are not appreciably shunted by the valve and stray capacitances. This usually involves the use of resistors of a few thousand ohms only, thereby again limiting the gain per stage.

It is important to note that both at the low and high ends of the scale the consideration is not so much the falling off in the response as the phase shift. If the phase of the output current is not in the same relationship as on the input, then distortion will appear on the picture. For example, the highest frequencies are caused by sharp changes from black to white. If we have a fine check pattern somewhere in the

image, then we shall get a succession of very high frequency currents every time the scanning reaches this point. If there is any phase displacement in the amplifiers, the effect will be to retard the response from the receiver so that the check pattern on the receiver will be displaced from its normal position.

Phase displacement occurs much earlier than any noticeable falling off in the gain, so that the requirements are far more stringent for a television amplifier than for a radio amplifier.

It is easy to show* that the low-frequency phase angle is $\tan^{-1}(1/RC\omega)$, R being the grid leak and C the capacitance of

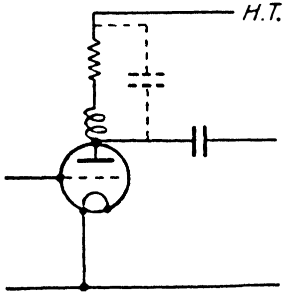

FIG. 83. LOSS OF HIGH FREQUENCIES IN A RESISTANCE-COUPLED AMPLIFIER MAY BE MINIMIZED BY USING AN INDUCTANCE IN SERIES WITH THE RESISTANCE

the coupling condenser. For a phase shift of 5° only, $1/RC\omega = 0{\cdot}0875$. At 25 cycles, $\omega = 157$, whence the product $RC = 0{\cdot}073$.

The high-frequency phase shift is $\tan^{-1} RC_0\omega$, R in this case being the anode resistance and C_0 the total stray capacitance, including valve capacitances, across the circuit. For 5° shift we have $RC_0\omega = 0{\cdot}0875$. If the highest modulation frequency is 2 Mc/s., $\omega = 12{\cdot}56 \times 10^6$ and the product $RC_0 = 28$, where R is in thousands of ohms and C_0 is in micro-microfarads. If $R = 1\,000$ ohms, C_0 can only be $7\mu\mu$F. Even so, the gain with a valve having a conductance of 4 would only be 4 per stage.

In practice, a slightly greater phase shift has to be tolerated,

* See *Modern Radio Communication*, Vol. II, Chap. VII.

and it can be taken that if $RC_0 = 40$ to 50 kilohm-micro-microfarads, satisfactory reception will result. Small inductances are sometimes introduced in series with the resistance, as shown in Fig. 83, to provide a resonance with the stray capacitance and maintain the response for a little longer.

For further information, refer to "Television Receiving Circuits," by Robinson, *Proc. I.R.E.*, June, 1933, and also to *Television, Theory and Practice*, by the author (published by Chapman & Hall).

CHAPTER X

FREQUENCY MODULATION

It is usually easier to vary the frequency of an oscillator than to change the amplitude, and various attempts have been made to utilize this fact. In particular, keying is sometimes accomplished by changing from the "marking" wave to a "spacing" wave on a slightly different frequency, just far enough away for the receiver to discriminate satisfactorily.

Attempts to use frequency modulation for telephonic and other forms of communication were, for a long time, not so successful, but methods have been evolved in the past few years which are making considerable headway.

Two possibilities exist. We can—

(*a*) Vary the carrier frequency by an amount equal to the modulation frequency desired.

(*b*) Vary the carrier by a given percentage *at a rate* depending on the modulation frequency desired.

Although the first method seems the most straightforward, it is the second which is usually employed in practice. The carrier is varied in frequency a few per cent on either side of the normal at a rate depending on the modulation. Thus for a 500-cycle tone, the carrier would go through the frequency cycle $f, f + \delta f, f, f - \delta f, f$ in 1/500 sec., while for a 2 000-cycle tone it would do the same cycle in 1/2 000 sec.

The side-bands produced by a frequency modulated wave are very complex and depend on the relation between the variation δf and the modulation frequency. For further information, refer to "Frequency Modulation," by Van der Pol, *Proc. I.R.E.*, July, 1930; and "Amplitude, Phase and Frequency Modulation," by Roder, *Proc. I.R.E.*, Dec., 1931.

Frequency modulation was first used in commercial practice for providing a tone on a c.w. telegraph signal, to obtain a band spread, and thereby minimize selective fading. It is better than amplitude modulation in this respect since it does not involve extra power, the carrier amplitude being constant

and the modulation being produced by a shifting of phase rather in the fashion of Fig. 59.

Armstrong Frequency Modulation System.

Towards the end of 1935, frequency modulated telephony received a considerable impetus from the publication of a paper by E. H. Armstrong, the originator of the super-regenerative circuit.* This paper disclosed details of a system

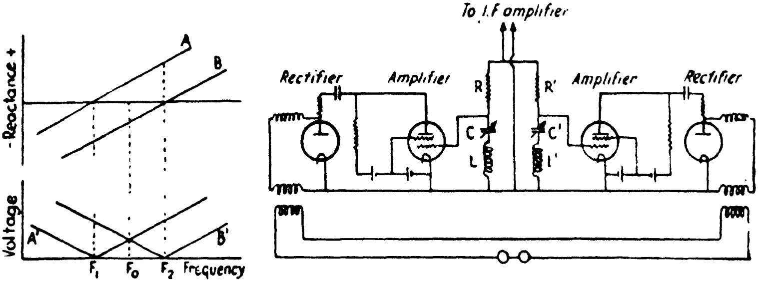

FIG. 84. ARMSTRONG FREQUENCY-MODULATED RECEIVER CIRCUIT

which had been undergoing trials for some two years on a frequency of 41 Mc/s. with remarkable results, one of the most striking being the very marked improvement in the signal-to-noise ratio. On several occasions of bad electrical conditions, frequency-modulated signals transmitted with a few hundred watts of power over a distance of 85 miles were received better than amplitude-modulated signals from a 50 kw. station 20 miles away.

The method of transmission was similar to that just discussed, the frequency being changed by an amount proportional to the amplitude of the modulation. Thus for full signal the final frequency was changed by $\pm$ 75 kc/s. while for no signal there was no change of frequency. The rate of change of frequency was dependent on the modulation frequency, being actually inversely proportional to the modulation frequency.

* E. H. Armstrong. "Reducing Disturbances in Radio Signalling by a System of Frequency Modulation." *Proc. I.R.E.*, May 1936, p. 689.

At the receiver the frequency modulation was converted to amplitude modulation by an ingenious arrangement shown, in essence, in Fig. 84. The frequency-modulated carrier was amplified by two successive intermediate frequency (superheterodyne) amplifiers operating at 6 and 0·4 Mc/s. respectively. These were wide-band amplifiers accepting a frequency spread of 150 kc/s. (made necessary by the $\pm$ 75 kc/s. deviation of carrier frequency), and the output therefrom was applied across two series-tuned circuits RLC and $R'L'C'$ in parallel. L and C resonate at the lower limit of the band (3 925 kc/s.), while L' and C' resonate at the upper limit (4 075 kc/s.). R and R' are swamping resistors to make the current substantially constant over the whole frequency range. The reactances of the two branches vary directly with frequency, as at A and B in the left-hand diagram, and because of the constant current through the branches, the voltage across CL or $C'L'$ varies linearly with frequency, as at A' and B', being zero at the tuning points and rising proportionally on either side.

Hence, as the frequency varies on either side of the carrier, the voltage across the two branches rises and falls in push-pull fashion and in direct proportion thus converting the frequency deviations into a change in amplitude. These amplitude variations will take place at a frequency determined by the rate of the frequency deviations in the original carrier. The result is thus a normal amplitude-modulated wave, which is amplified, rectified, and combined in push-pull to give the audio-frequency output.

The improved signal-noise ratio arises from the fact that extraneous noise picked up on an aerial is mainly amplitude modulated, whereas the receiver only responds to frequency modulation, and later developments have fully justified the original claims.

Methods of Frequency Modulation.

Since 1935 frequency modulation has made rapid strides. There are already a number of frequency modulated broadcast stations operating in America, and special adaptors are supplied to enable existing receivers to be used with this modified form of transmission. Commercial frequency modulated services are in operation, the receivers in this case, of

course, being specifically designed for the purpose. A detailed analysis of the process of frequency modulation involves mathematical treatment, but fortunately it is not necessary to discuss this aspect of the question in great detail.

Fig. 85 shows a skeleton frequency modulated transmitter. It consists of a simple oscillating valve comprising the oscillating circuit *LCR* which is maintained in a state of oscillation by means of a reaction coil in the anode circuit. Across the whole circuit is the control valve V_1, the grid of which is

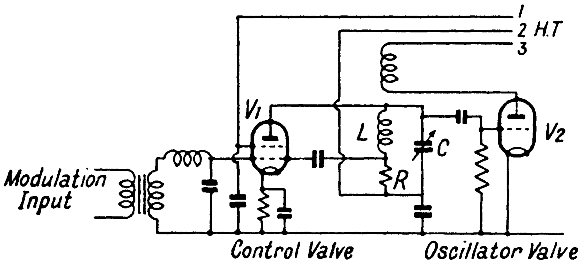

FIG. 85. F-M OSCILLATOR USING REACTANCE CONTROL VALVE

fed with voltage developed across the resistance *R*. Now the voltage across a tuned circuit is 90° out of phase with the oscillating current itself, but the voltage across *R* is in phase with the current. Consequently the voltage applied to the grid of V_1 is 90° out of phase with that across the tuned circuit, but since the anode current variation in V_1 will be in exactly opposite phase with the voltage on the grid (because the anode circuit contains the parallel turned circuit *LCR* which looks like a pure resistance) it follows that it is also 90° out of phase with the voltage across the tuned circuit.

But a deviee in which the current is 90° out of phase with the voltage has the characteristics of a pure reactance so that the control valve V_1 looks like a reactance (actually a capacitance) across the tuned circuit and the frequency will be modified accordingly.

Moreover, the extent of this reactance depends upon the value of the anode current in V_1 which is determined by the amplification of this valve. If this is of a vari-mu type the amplification will vary according to the bias on the grid.

The valve therefore is provided with a steady bias and its instantaneous value is varied by applying the modulation frequency required. Consequently, the frequency of the oscillation generated by V_2 will depend upon the audio-frequency voltage applied to the grid of V_1.

The greater the audio-frequency swing on the grid of V_1, the greater will be the change in frequency of the oscillation generated by V_2, and by suitably designing the circuit it is possible to arrange that with full modulation the change in frequency produced is equal to the deviation specified (say, $\pm$ 75 kc/s. in 40 Mc/s.). V_1 must be arranged so that over this range of operation its action is linear so that the frequency deviation produced is strictly proportional to the amplitude of the applied audio-frequency voltage. The rate at which the deviations occur will obviously depend upon the frequency of the modulation, being relatively slow at low modulation frequencies and rapid at high frequencies. Thus we produce a frequency modulated carrier having the characteristics already laid down.

There are various alternative methods of producing frequency modulation, though the ones commonly employed all utilize a variant of this reactance control arrangement. For this to function successfully it is necessary that the radio-frequency voltage applied to the grid should be accurately 90° out of phase with the voltage across the oscillating circuit, and there are various alternative ways of providing this out-of-phase control voltage which need not be described in detail. The frequency modulated output from V_2 would be passed through amplifier stages and ultimately to the aerial. In addition it is often convenient to arrange the primary oscillator at a lower frequency and to pass the output through frequency multiplying stages in accordance with the technique outlined in Chapter VI. In such a case the actual frequency deviation of the primary oscillator will be suitably proportioned so that at the end of the final multiplier it occupies the allotted spectrum of $\pm$ 75 kc/s. or whatever has been chosen.

Stability.

Now this simple form of circuit possesses one serious disadvantage in that it is not easy to control accurately the

mean carrier frequency about which the deviation takes place. Indeed an oscillator which is to respond to some control device which varies in frequency is clearly not likely of itself to be very stable in frequency.

High stability of the carrier might not appear to be of great importance, but it is a peculiarity of a frequency modulated system that a substantial part of the total energy lies in the extreme side-bands so that any wandering of the mean frequency would cause considerable interference between one channel and the next.

Wandering of the carrier frequency is also liable to give rise to asymetrical modulation, while if the modulation picked up by the receiver extends beyond the band width for which it has been designed much more distortion occurs than is the case with amplitude modulated arrangements. Because of these various factors the American Federal Communications Commission requires the carrier frequency to be constant within $\pm$ 2 000 c/s., which at a carrier frequency of 40 Mc/s. requires a stability of ·05 per cent.

Phase Modulation.

Because of this difficulty Armstrong in his original experiments utilized phase modulation; to understand this it is desirable to refer back to the discussion on modulation in Chapter VII (pages 116 and 117). It was shown here that with ordinary amplitude modulation the amplitude of the carrier would vary in the required manner only so long as the carrier and side-bands remained in phase, but that if the carrier is out of phase the effective depth of modulation is reduced until with a 90° phase shift the amplitude modulation becomes very small and is replaced by a change in phase of the carrier.

Referring back to Fig. 59 it will be seen that during progressive stages of the modulation cycle the resultant wave first lags behind the original carrier and then gains until it is in phase again, after which it begins to lead and finally falls back into phase again. Consequently the phase of the wave is continually swinging behind and ahead of the original carrier and such a wave is said to be phase modulated. It

will be clear that such a wave must at the same time be frequency modulated because for the phase to fall behind the frequency must temporarily be reduced and vice versa, and mathematically it can be shown that frequency modulation is equivalent to a phase modulation in which the phase shift is inversely proportional to the modulation frequency.

It will also be clear that in order to convert amplitude modulation into phase modulation it is only necessary to shift the phase of the carrier by 90° and Armstrong's method, which has been adopted by many other experimenters, was to provide a stable crystal controlled oscillation which was amplitude modulated in accordance with known technique to a degree which is inversely proportional to the signal frequency. The carrier wave was then separated from the side-bands, shifted 90° in phase and then re-combined with the side-bands.

This arrangement has the advantage that the mean carrier frequency can be held constant to a high degree of stability. It suffers from one serious disadvantage in that if distortionless modulation is to be produced the phase shift should not exceed about 60° either way, whereas in order to produce the modulation required at low frequencies a phase shift of several thousand degrees is necessary. This disadvantage is overcome at the expense of considerable complexity by a very high degree of frequency multiplication amounting to several thousand times. The operation is indeed sometimes carried out in two stages, the frequency being multiplied a large number of times and then being made to beat with another oscillator (also of stable characteristics) to produce a very low mean frequency in accordance with the ordinary beat frequency principles. Since the operation depends on frequency differences, however, the phase shift remains unaltered by this operation, and subsequent multiplications may then be resorted to bringing the frequency up to its original value, while the phase shift is progressively increased until the required phase shift is produced.

Automatic Carrier Stabilization.

An alternative arrangement was described by Morrison—*Proc. I.R.E.*, October, 1940—in which an oscillator of the

type of Fig. 85 was used, but the carrier was kept constant by a form of automatic tuning control. It is a peculiarity of the frequency modulated system that the side-band amplitudes are considerably greater than the carrier, so that the selection of the carrier to operate any such control is difficult. Morrison, therefore, generates his frequency at approximately 5 kc/s. and frequency modulates it by a reactance control valve. The output is passed through a buffer amplifier and is then split. Part of it goes through frequency multipliers providing a multiplication of eight times to bring the frequency up to the 40 Mc/s. required, while the other part goes through frequency dividing stages until it is reduced to a frequency of the order of one kc/s. As a result of this the upper side-bands are progressively lost and the carrier is proportionately much greater, so that it can be filtered out and applied to a discriminator circuit of the type which is common with automatic tuning control. This carrier is then referred against a standard constant frequency, and any deviation from this standard value operates through the discriminator to rotate a tuning motor in one direction or the other to restore the original frequency to the required amount.

Because of the development taking place detailed transmitter circuits are not frequently encountered, though American literature is the best source of this information at the present time. The fundamental theory already outlined, however, should enable the reader to follow any such literature without difficulty.

Frequency Modulated Receivers.

The reception of frequency modulated signals involves some form of discriminator circuit as already indicated in Fig. 84. The initial stages of the receiver are similar to those of an amplitude modulated instrument consisting of an r.f. amplifier, frequency changer and i.f. amplifier. At this point the signals are passed through a limiter, the object of which is to maintain the amplitude constant irrespective of any fluctuations in amplitude of the received signal, whether these be originally present or arise, as is more likely, from atmospheric and/or locally generated disturbances (e.g. motor car ignition systems).

The essential feature of a frequency modulated transmission

is that its amplitude is constant, and it is thus permissible to smooth out any such variations completely. If then the detector circuit is so designed that it will not respond to any changes in amplitude, all extraneous noise is eliminated, leaving only that generated by the receiver itself. Fig. 86 illustrates such a limiter and its succeeding detector. The limiter is an r.f. pentode in which the anode voltage is cut down to about 20 volts. Under these conditions a saturation effect is produced, and beyond a certain critical value the

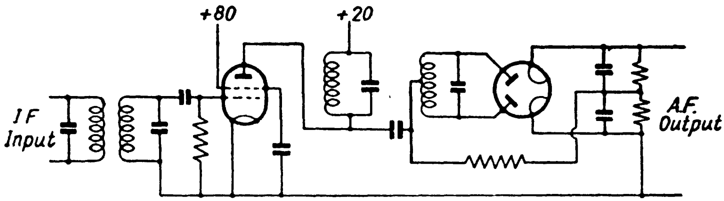

FIG. 86. DETECTOR SYSTEM FOR F-M RECEIVER

grid voltage can be varied quite widely without producing any change in anode current.

The output from this stage is then passed through the discriminator circuit. In the present instance this operates on a slightly different principle, one which is frequently used for automatic frequency control. It depends upon the fact that the voltage in the primary and secondary circuits of a loosely coupled tuned transformer are 90° out of phase. If these two voltages are added together the maximum response will not occur at the frequency of tune, being of the form shown in Fig. 87 (*a*). If the direction of secondary is reversed the response curve is as shown by the dotted line.

If we can arrange to subtract the voltages from one another we shall in effect invert the dotted line, and the resultant is then of the form shown in Fig. 87 (*b*). This will be seen to have a substantially straight portion over which the change in frequency produces a strictly proportional change in amplitude.

The subtraction of the voltages is, of course, arranged by suitably choosing the directions of the windings, and it will be noted that deviations in frequency in either direction will produce a change in amplitude in the appropriate sense. The

primary is, therefore, connected through a suitable isolating condenser to the mid-point of the secondary, and a push-pull detector is arranged so that both positive and negative deviations (i.e. both side-bands) are utilized. The audio-frequency output is then applied to the output stage in the

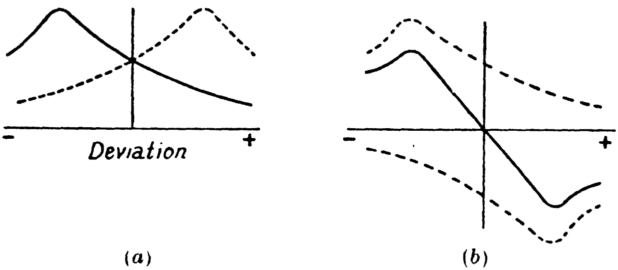

FIG. 87. ILLUSTRATING OPERATION OF DISCRIMINATOR

customary manner. A more detailed description of a typical frequency modulated receiver appeared in *Wireless World*, December, 1940.

Other Applications.

Apart from the marked advantages in signal to noise ratio which results from frequency modulation, there are various other applications which are only just beginning to be exploited. One of these is in the more accurate measurement of distances by reflection, as for example in the determination of the actual height of an aircraft. (The ordinary altimeter, of course, operates on barometric principles and only indicates the height above given reference level.)

Such an arrangement was described by Espenschied and Newhouse in the *Bell System Technical Journal*, January, 1939. In brief, a frequency modulated wave is sent to the ground and the time is noted for its return to the aircraft. This time is accurately determined by noting the change in frequency which has taken place, and since the frequency is changing at a known rate the time can be determined with precision.

Numerous other applications are coming to the fore, and it is hardly too much to say that the successful development of frequency modulating technique is opening out an entirely new field of operations for the radio engineer.

CHAPTER XI

MICRO-WAVES

IT has already been explained that the ordinary methods of generating oscillations with a valve break down at frequencies of the order of 300 Mc/s., corresponding to a wavelength of 1 metre. This is because the electrons take a finite time to traverse the gap between the electrodes, and the input conductance of the valve becomes increasingly large.

This, together with the limited range of the ultra-short waves, produced a slowing down in the triumphal progress of the engineers. But there were gaps in the curtain through which further progress could be envisaged. These gaps were explored, this time not by the amateurs (for indeed a highly skilled technique was required), but by the physicists, into whose domain the radio engineer was rapidly encroaching.

The valve engineer was the first to move. New types of valve were evolved, requiring new forms of circuit. Meanwhile the physicists were attacking the problems of transmission. Theories propounded as long ago as 1893 became susceptible to practical examination due to improved technique, and an entirely fresh field of endeavour has been opened up, covering frequencies ranging from 300 to 300 000 Mc/s. (1 m. to 1 mm.). Actually radio waves as short as 0·1 mm. have been produced by Arkadiewa in Russia, but below 1 mm. we merge into the region of infra-red (heat) waves which can be produced by physical rather than radio technique.

These waves from 1 m. to 1 mm. are known as *micro-waves*. They cover a spectrum ten times as wide as that occupied by the entire short wave field and, as mentioned in Chapter I, will soon require a literature of their own to keep pace with the new and radically different technique which is already appearing.

It is only possible in the present work to indicate the trend of progress, and thus provide the reader with a link between orthodox short-wave practice and the new technique. The

present chapter deals with the work of the valve engineer, while the new forms of transmission are considered in Chapter XII.

Barkhausen-Kurz Oscillations.

In 1920 two German engineers, Barkhausen and Kurz, produced oscillations, using a circuit of the type shown in Fig. 88, at frequencies much higher than anything previously achieved. The grid was run at a positive potential with little or no volts

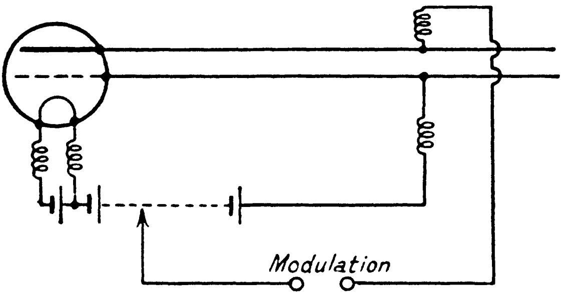

FIG. 88. BARKHAUSEN-KURZ CIRCUIT

on the anode, and the explanation put forward was that electrons leaving the cathode were first attracted to the grid, shot through the spaces therein, and were brought to rest by the retarding field between grid and anode due to the lower potential of the latter electrode. They then commenced to return (due to the positive grid potential), shot through the grid again, and were once more brought to rest by the retarding field in the grid-cathode space. They then reversed again, and so an oscillation was built up independent of the external circuit and having a frequency dependent on the dimensions and operating voltages of the valve.

Such an arrangement is known as a *B–K oscillator*, and in the original experiments wavelengths ranging from 43 to 200 cm. were generated, the actual wavelength being given by

$$\lambda\,(\text{cm.}) = \frac{2\,000}{\sqrt{V_g}} \cdot \frac{AV_g - BV_a}{V_g - V_a}$$

where V_g and V_a are the grid and anode voltages, A is the distance from anode to filament, and B the distance from grid to filament (both in cm.).

If $V_a = 0$, this reduces to the form $\lambda^2 V_g = k$, k being a constant.

The oscillation is extracted from the valve by connecting a tuned feeder or *Lecher wire* to grid and anode as in Fig. 88. The necessary d.c. voltages are fed in at nodal points, with h.f. chokes in the leads. In the original circuit, filament chokes were also used.

Gill-Morrell Oscillations.

In 1922, Gill and Morrell, working with a similar circuit, but using a valve having a greater anode-grid separation, obtained oscillations covering a range up to 500 cm. They found, however, that the oscillation was *not* independent of the external circuit, but varied somewhat as might be expected. At certain points, however, the oscillation was much more intense, and the wavelength at such points was related to the grid potential by an expression similar to that just quoted. Actually, Gill gives the formula $\lambda^2 I_g / \sqrt{V_g} = k$, where I_g is the grid current, which reduces to the B–K formula for zero anode voltage if I_g varies as $V_g^{3/2}$. Oscillations of this type are usually termed *G–M oscillations*.

The dependence of the oscillation on the external circuit is, of course, not accounted for by the simple B–K theory, and for many years it was considered that there were two essentially different modes of oscillation. In 1936, however, Anderson pointed out in *Electronics* (August, 1936) that the Fig. 88 circuit is, in effect, an ultraudion oscillator, and that if such an oscillator is analysed under conditions of low grid-cathode resistance it can be shown that there is a limiting frequency below which the circuit will not maintain oscillation.

This limit is given by the expression $\omega_0^2 = 1/C_a C_g r_a r_g$, where C_a and C_g are the anode-cathode and grid-cathode capacitances, and r_a and r_g are the corresponding internal a.c. resistances.

Normally this limiting frequency is well below that determined by the external circuit, but if r_g is low (as it will be if

the grid is positive), the critical frequency may be above that of the external circuit.

Transference of Power.

Later investigations appear to support the suggestion that there is no essential difference between the two forms of oscillation, but that the external circuit may have more effect in certain circumstances. This is made clearer if we examine more closely the means by which power is transferred from the valve to the external circuit.

Let us assume that we have a circuit connected across grid and anode as in Fig. 88, and that an oscillation exists in this circuit of a frequency comparable with the natural frequency of electron oscillation within the valve. The grid is normally positive, but will fluctuate above or below its mean value at the frequency of the (external) oscillation.

When the grid is more positive than normal it will cause the electrons leaving the cathode to accelerate more rapidly; but this requires additional energy which will be extracted from the external circuit. Conversely, when the grid is less positive than normal the electrons will be retarded, and will give up energy to the external circuit.

These two conditions would normally balance each other but for the fact that electrons which are accelerated more rapidly than normal cannot stop before they reach the anode, which therefore collects them and drains them from the system. The retarded electrons, on the other hand, remain in the system and give up some part of their energy.

It will be clear that this action can take place even if the frequency of the external circuit is not the same as the natural frequency of the electrons within the valve, so that it is possible for the frequency to be controlled by the external circuit. Clearly, however, the results are best when the external and internal frequencies agree, or are in simple harmonic relation. Good results are obtainable, for example, if the external circuit is tuned to twice the frequency of the valve, and this is often done.

The energy available is small, so that satisfactory operation is only possible with external circuits of high Q, which is the

reason for the use of the tuned feeder which is the most efficient form of tuned circuit at these frequencies.

It will also be clear that the energy dissipated by the accelerated electrons is of the same order as the energy delivered by the retarded electrons. But the accelerated electrons are collected by the anode, so that anode current only flows when oscillations are occurring externally, and the amount of anode current is a measure of the amplitude of the external oscillations.

Practical Forms of Circuit.

To increase the power, several valves may be connected in parallel to the same feeder, and this may be led to a di-pole

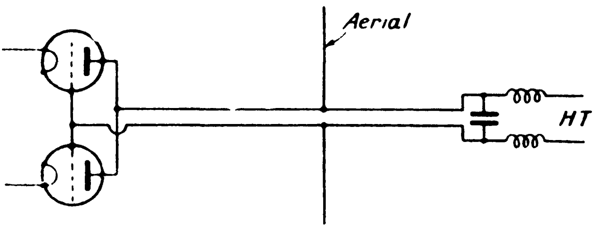

FIG. 89. TWO ELECTRON OSCILLATORS IN PARALLEL

aerial situated at a suitable node. Two valves in parallel are illustrated in Fig. 89, together with an aerial system.

An alternative arrangement due to Mathieu is shown in Fig. 90. Here, two valves are used in a sort of push-pull arrangement. The feeder wire is connected to the two grids, and not to grid and anode as in the previous arrangement, while modulation can be obtained if desired by varying the anode potential with a modulation transformer in series with the h.t. supply. The cathodes and the earth point are connected via resonant lines adjusted in length so that the batteries or earth connection occur at voltage nodes (points of zero potential).

A feature of this arrangement is its symmetry, and the circuit has been used extensively by the Marconi Co. Quite often groups of oscillators of this type are employed to feed aerial arrays similar to, but on a smaller scale than, those

already mentioned in Chapter IV. The transmitters in this case become an integral part of the array itself.

It is impossible to discuss the matter in more detail here, and for further information the reader is recommended to refer to a paper by Megaw, *Journal I.E.E.*, Vol. 72, p. 313: and also to a description by McPherson and Ullrich of the Lympne-St. Inglevert Micro-ray Link, *Journal I.E.E.*, Vol. 78, p. 629.

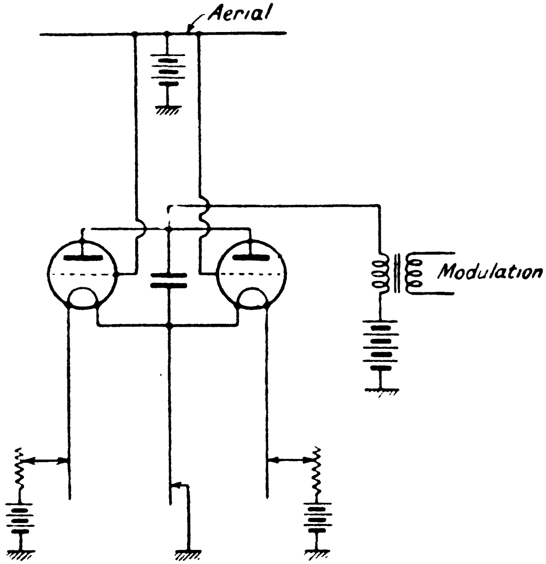

FIG. 90. MATHIEU MICRO-WAVE OSCILLATOR

It is worth noting that in the system described in the last-named paper the method of extracting the energy is slightly different in that the external circuit is connected to the two ends of the grid (in which the usual longitudinal support wires are omitted so that it forms a small inductance). This form of circuit has been investigated by various writers, but in general terms the action can be considered as similar to that just described, one end of the grid being more positive than normal, and the other more negative, at any given instant. The circuit is illustrated in Fig. 99.

The Magnetron.

Another form of generator suitable for very high frequencies is the magnetron, which is an arrangement comprising a cathode and a cylindrical anode, with a magnetic field applied

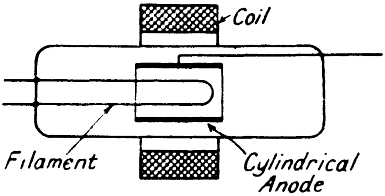

Fig. 91. Simple Magnetron Valve

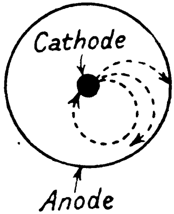

Fig. 92. Illustrating Action of Magnetron

axially down the tube by means of an external coil, as shown in Fig. 91.

Now, the effect of a magnetic field on the electron stream is curious. Electrons leave the cathode in all directions in radial straight lines. The magnetic field, however, causes them to deviate to one side, so that they proceed in a curved path as indicated in Fig. 92.

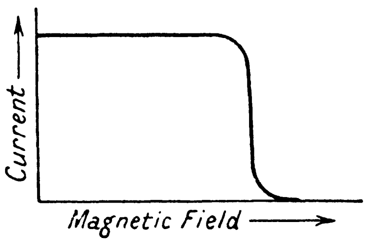

Fig. 93. Magnetron Characteristic

As the strength of the field increases, the curvature of the path becomes more and more until a critical value is reached, at which the electrons fail to reach the anode altogether and return to the cathode region. As they do so, of course, they will be slowed down, and will ultimately come to rest and start off again. The net effect, however, will be that no current actually reaches the anode, and there will be quite a sharp dividing line between the condition of no anode current and the condition of normal current. Up to this critical value the magnetic field has little or no influence on the actual current, as will be clear from the

explanation just given. A typical magnetron characteristic is shown in Fig. 93.

Electron Oscillations with the Magnetron.

It is clear that the electrons which leave the cathode and do not reach the anode must be describing roughly circular orbits in a definite time so that within the valve itself there are electronic oscillations. If we consider a valve operating around the cut-off region it is clear that the actual anode current can be varied by altering the potential of the anode over a small amount. We are therefore able to extract energy from the valve by a process rather similar to that already described for the positive grid electron oscillator.

If the anode potential decreases, the electrons in the vicinity will be retarded and therefore energy will be extracted from the system, while if the anode potential is increased the electrons will be accelerated and caused to reach the anode, where they will be drained out of the circuit, leaving us with a net acquisition of energy.

It is again necessary that the variation of anode potential shall be suitably timed relative to the natural period of circulation of the electrons within the valve, so that the frequency of oscillations of this type is determined by the valve itself, but with the important advantage that it is possible to alter this natural frequency by altering the value of magnetic field, whereas with the positive grid type of oscillator the natural frequency of oscillation is determined by the physical dimensions of the valve.

It is found, in practice, to be most convenient to use an arrangement in which the anode is divided into two halves which are connected in push-pull fashion as indicated in Fig. 94. The external circuit is again made in the form of a resonant line, the length being adjusted to be half a wavelength, and under such conditions the wavelength of the oscillation produced is given approximately by

$$\lambda = 11\,000/H \text{ cm.}$$

where H is the field strength in gauss.

This is an empirical expression based on calculation of the transit time of the electron in its orbit from cathode back to

cathode again. It will be noticed that it only depends upon the strength of the magnetic field, but, remembering that it is necessary to operate the valve around the cut-off condition, it will be clear that this, at the same time, necessitates a certain value for the anode potential, this value being determined by the physical structure of the valve. Clearly also, up to a point, the higher the anode potential the more the energy which can be extracted from the valve.

This treatment is necessarily very brief. A great deal of work has been done on this class of valve, and it is actually with this type of circuit that the highest frequencies have

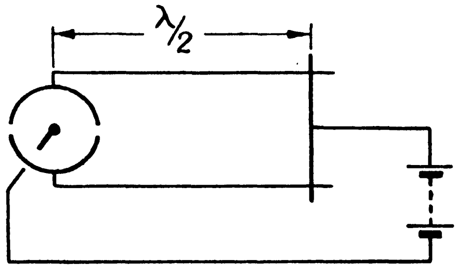

FIG. 94. SPLIT ANODE MAGNETRON CIRCUIT

been generated. The references given at the end of this chapter will provide the reader with the opportunity of studying the subject further.

Dynatron Oscillations.

The amount of power extracted from the electronic form of magnetron oscillation is small. Efficiencies of a few per cent only are quite normal. At longer wavelengths, however, it is possible to use the valve in a considerably more efficient manner by making use of a negative resistance effect which can be produced with the split-anode type of valve.

This arises from the change in the configuration of the electric field around the gap. Fig. 95 (*a*) shows the condition with both anodes at the same potential. The force on an electron emitted from the cathode is two-fold. The electric force is along the line of electric field at the point concerned, while the magnetic force is always *at right angles to the direction*

of motion of the electron. The electron starts off radially but immediately becomes subjected to a force at right angles, which causes it to deviate from its initial course. At each point in the interelectrode space the effective force on the electron is thus the resultant of a radial force and a force at right angles to its motion. These forces are indicated at various stages in Fig. 95 (*a*), and result in a spiral motion as shown.

If the magnetic field is insufficient the electron strikes the anode. As the field strength is increased the path of the electron just misses the anode, and returns to the cathode as in Fig. 92. With still higher values of field the electron does not reach the cathode on its first circuit, but spirals round the

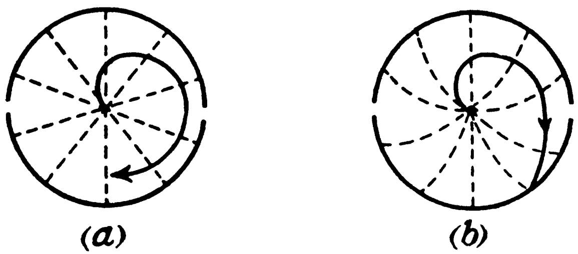

FIG. 95. ILLUSTRATING DYNATRON ACTION WITH SPLIT-ANODE MAGNETRON

cathode two or three times, thus having a longer transit time and generating a longer wavelength.

This is the mechanism of the electronic oscillation just discussed. If, however, we have an appreciable difference of potential between the two sections of the anode, the conditions become as shown in Fig. 95 (*b*), where the top section is more positive than the bottom. It will be seen that the electric field is no longer radial in the region around the gap, and as a result the deflection of the electron is diverted so that it does not continue to spiral inwards but strikes the bottom plate. Thus reducing the potential of the bottom anode has produced an *increase* of current, so that the valve exhibits a negative resistance.

By connecting the valve as shown in Fig. 96, therefore, the external circuit may be maintained in a state of oscillation. The frequency of such oscillation is controlled by the external

circuit since it depends on the dynatron action just discussed and not on internal electronic movements, but the valve will only maintain this class of oscillation at relatively long wavelengths, such that the period of oscillation is long compared with the transit time.

As the wavelength is reduced the efficiency falls from about 50 per cent (at frequencies around 100 Mc/s), becoming progressively lower until the oscillations merge into the electronic type and cease to be controlled by the external circuit.

Deflection Oscillators.

The development of the cathode ray tube has led to attempts to use this device as a high frequency generator. As the

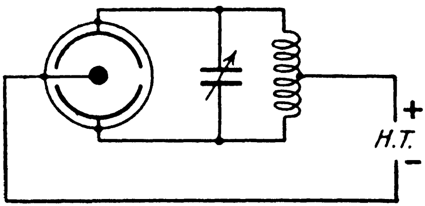

FIG. 96. DYNATRON MAGNETRON CIRCUIT

reader will be aware, the cathode ray tube is a device in which a stream of electrons is produced at one end of a long tube and projected down the tube to a fluorescent screen at the far end. The beam may be deflected from its normal position by arranging plates on each side of the beam, near its point of origin, and applying a potential difference across the plates.

If we replace the fluorescent screen with two electrodes the beam may be deflected, by applying a high frequency oscillating voltage to the deflector plates, so that it falls alternately on each of the collector electrodes, and if these are connected to the opposite ends of a tuned circuit or resonant line the device can act as an amplifier or, with suitable feedback, as an oscillator.

The arrangement is only partially successful because the beam has a high resistance which limits the power handling capacity while it is necessary to have an output circuit of high

impedance, which is again difficult at really high frequencies. The arrangement is discussed further in Chapter XIII (Fig. 115).

Velocity Modulated Oscillators.

The transit time of the electrons is a further element to be contended with. In 1936, however, details appeared of a new type of generator in which, instead of trying to minimize these defects, they were utilized to produce a radically different form of device. The problem of obtaining a high impedance was overcome by the use of resonance cavities built into the generator.

Consideration of the latest class of oscillator will, therefore, be deferred until we have examined the matter of resonance chambers and wave guides, since these in themselves constitute a novel technique, at any rate for the engineer. Further details of velocity-modulated oscillators will be found in Chapter XIII.

REFERENCES

"Electronic Oscillations." Megaw. *Journal I.E.E.*, Vol. 72, April, 1933.

"Micro-Ray Communication." McPherson and Ullrich. *Journal I.E.E.*, Vol. 78, June, 1936.

"The Magnetron Short-wave Oscillator." Megaw. *Journal I.E.E.*, Vol. 72, April, 1933.

"Magnetron Oscillators." Kilgore. *Proc. I.R.E.*, Vol. 24, August, 1936.

"Anode Tank-circuit Magnetron." Linder. *Proc. I.R.E.*, Vol. 27, January, 1939.

"Resonance and Electronic Oscillations in Magnetrons." McPetrie and Ford. *Journal I.E.E.*, Vol. 86, March, 1940.

CHAPTER XII

TRANSMISSION AND PROPAGATION OF MICRO-WAVES

ONE great advantage of micro-waves is that the wavelength is short enough to permit of solid reflectors being used as in the case of light waves. This is more particularly the case with the shorter micro-waves below about 50 cm., while for the larger waves the technique tends more to follow the arrangement of aerial arrays employed with normal short-wave working. In between the two a combination of the two methods may be employed, one particular form being illustrated in Fig. 97. This will be seen to consist of a number of resonators half a wavelength long spaced along a parabolic backbone.

Any array, however, has the disadvantage that it only applies for one wavelength, whereas if a solid reflector can be used it applies equally well to any wavelength reasonably short compared with the dimensions of the reflector. If the reflector can be made to have an aperture (i.e. an effective diameter) of some 8 to 10 wavelengths, this is found to apply.

A parabolic metal reflector will concentrate the rays falling on it into a parallel beam, exactly as in optics, provided the radiator is arranged at the focal point. The reflector, however, only operates on the waves radiated *back* from the aerial to the reflector. The forward radiation still spreads out in all directions and is thus largely wasted. To overcome this, a spherical mirror may be used in front of the aerial. This reflects the waves back through the focal point, as shown in Fig. 98, so that they ultimately reach the parabolic reflector and come back within the beam. OAB is a backward wave reflected into a parallel beam in the normal way. The other way is a forward one running from O to C, back to O, on to D, and then out to E.

Lenses have also been built of ebonite for the concentration of these micro-waves. In particular, on the English Channel link, experiments were made with a lens 70 cm. in diameter

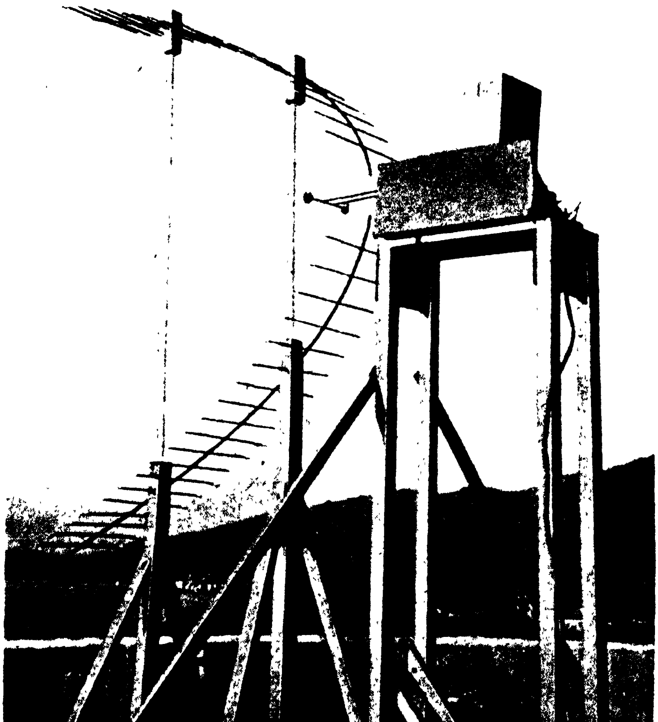

FIG. 97. MICRO-WAVE TRANSMITTER
(Marconi's Wireless Telegraph Co.)

placed in front of a half-wave vertical aerial excited at a wave-length of 19 cm. Definite concentration was obtained at a distance of 40 cm., the signal strength without the lens falling to approximately to one-tenth of its former value.

Various forms of reflector and zone plate are described in the paper by McPherson and Ullrich, referred to in the last chapter.

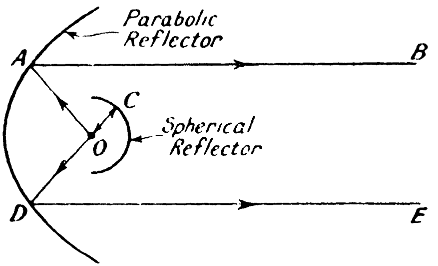

FIG. 98. REFLECTOR SYSTEM FOR MICRO-WAVES

Micro-wave Receivers.

Not the least difficult of the problems encountered in micro-wave work is that of the receiver. The transmitted wave-length is not by any means constant, so that a highly selective arrangement is quite unsuitable. There is, of course, no need for any substantial selectivity from the transmission point of view, partly because the restricted range of transmission makes it unlikely that there will be, in a given locality, more than a very few transmissions capable of causing interference; and secondly, because of the extremely sharp directional effects which can be produced.

It is found, however, that even normal selectivity is excessive and does not give an adequate factor of safety to cope with the wandering of the transmitted wavelength, which is inevitable with present-day technique. Possibly in time this drifting will be cured, but for the moment receivers must be very unselective when considered by normal standards.

The super-regenerative circuit is very suitable for this class of reception. The theoretical discussion of this was given in Chapter VIII, and the very broad tuning obtained with this class of circuit is a decided advantage.

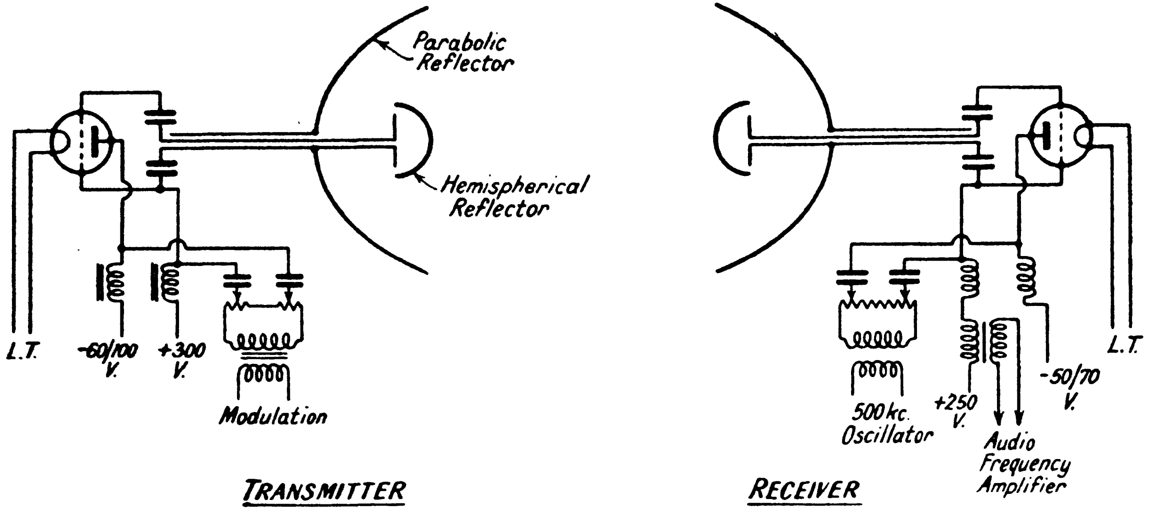

FIG. 99. DIAGRAM OF ENGLISH CHANNEL MICRO-WAVE LINK

A special form of electronic oscillator is used, connected by a tuned coaxial feeder to aerials housed in reflectors

The layout of the receiver usually resembles that of the transmitter. Fig. 99 shows the transmitting and receiving circuits of the Lympne micro-ray installation, operating on a wavelength of 17·4 cm., from which it will be seen that similar arrangements are used at both ends. We have seen that anode current appears in a *B–K* oscillator when electronic oscillations are present so that the valve in this respect is similar to an ordinary anode-bend detector. The circuit shown is of the super-regenerative type, the 500 kc/s. oscillator being for the purpose of quenching the oscillation.

Measuring Micro-waves.

The determination of the wavelength of micro-waves is best done by an actual measurement of the wavelength. By coupling to the circuit a tuned feeder, or Lecher wire, standing waves may be produced. The effective length of the feeder may be varied by connecting a short-circuiting bar across the wires. If this short-circuiting bar contains an indicating device, such as a flash lamp or vacuo-junction, it may be moved along the feeder until a maximum current is obtained, which will occur at a current node. By moving the bar along the feeder a succession of current nodes will occur, separated by half a wavelength, and the distance between them can be measured with exactitude. At the very high frequencies involved this is a much more accurate method than any attempt at frequency measurement.

Propagation of Micro-waves.

The optical qualities of micro-waves render them of considerable value in certain specified conditions. Atmospherics are absent, though the increasing use of very high frequencies in medical research is giving rise to some interference. Fading is only occasionally experienced, but does arise when belts of fog are settling in the path of the wave. Standing waves may be troublesome at the receiver, due to interference between the direct ray and one which has come nearly the same way and has been reflected from local obstacles.

There is no twisting of the plane of polarization, which is very useful; so much so, that the same reflector system can be used for two transmissions, one vertically polarized and

the other horizontally polarized. This is done on the English Channel link for two-way working, the outward and inward transmissions being polarized at right angles and transmitted or received with the same reflector.

Wave Guides.

Of recent years there has been a break away from the orthodox method of transmitting and radiating micro-waves. Before the close of the last century Lord Rayleigh had suggested that electric waves could be propagated through metal tubes without the usual go and return conductor mechanism. He even pointed out that it was possible for waves to be

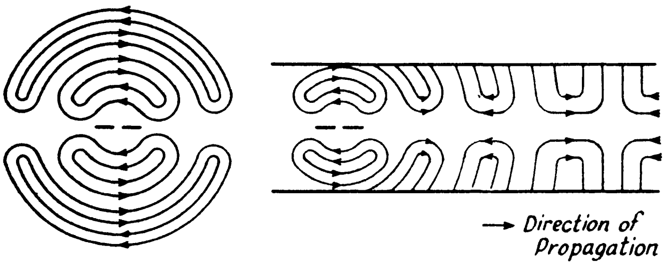

FIG. 100. ILLUSTRATING CONFINEMENT OF ELECTRIC FIELD WITHIN A GUIDE

transmitted through a cylinder consisting entirely of dielectric without any conductor at all. For over thirty years this possibility was never exploited, but in 1936 several investigators, notably Southworth and Barrow, operating independently, described experiments in the transmission of micro-waves through hollow conductors in the form of cylindrical or rectangular tubes.

In order to envisage this it is necessary to abandon the idea of current flowing along one conductor and returning along another, and to replace it by the conception of electric or magnetic fields. This is not so difficult for the radio engineer who is already accustomed to visualize the propagation of energy through the ether in the form of wireless waves.

Fig. 100 shows how this can be accomplished. At the left-hand side of the figure we have a dipole aerial and the oscillation of electrons to and fro in this aerial generates closed

loops of electric field which radiate outwards from the centre of the dipole. Now we know that if we enclose an electric field within a metal box, eddy currents are induced in the metal work which result in the production of equal and opposite field outside the box, consequent on which the effective field outside is zero. The effect is, in fact, the same as if the electric field had been constrained to flow within the material of the box. This is the usual principle of screening with which the reader will be familiar.

Let us see what happens if we enclose this dipole within a metal tube having its axis along the length of the dipole. This has been illustrated on the right-hand side of Fig. 100, and it will be seen that the effect is to restrict the outward propagation of the electric field in every direction except along the axis of the tube. We further see that a short distance away from the dipole the configuration of the electric field has settled down to a regular series of practically rectangular closed loops, and that these loops will travel along the tube by the same mechanism as waves travel in free space.

The arrangement shown would produce radiation in both directions along the tube, but it is clear that one could terminate the left-hand side of the tube in some manner which either absorbed the energy or reflected it in the correct phase, so that it strengthened the fields progressing towards the right, and in this case all the energy would be concentrated in the right-hand direction. This is the more usual arrangement, and in Fig. 100 the portion to the left of the dipole has been ignored.

It will be clear that for this device to be practicable it is necessary for the dipole to be small compared with the dimensions of the tube, so that it was not until it became practicable to generate wavelengths of centimetre order that this form of transmission came within the bounds of practice. But since we are now able to generate waves which are only a fraction of a centimetre in length it becomes clear that we can produce higher orders of wave. We could, for example, have two dipoles side by side fed with current 180° out of phase, thus producing fields having a configuration similar to that shown in Fig. 101.

It will be noted that the configuration of the electric field

has one important difference from that in a coaxial or parallel wire feeder. In the latter case the only field in the dielectric is at right angles to the direction of propagation. On the other hand, the field inside a wave guide will be seen to possess a definite axial component in the direction of propagation. In

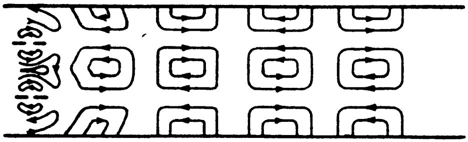

FIG. 101. HIGHER ORDER WAVES

the diagrams of Figs. 100 and 101 this axial component is electric, and the waves so produced are therefore known as E waves. Associated with these electric fields there are the usual magnetic fields just as in the case of the radiation of waves into space. The magnetic field in Fig. 100, for example, are a series of circular fields coaxial with the tube.

It is not difficult to appreciate that by rearranging the manner in which the wave is started in the guide it is possible to arrange longitudinal loops of *magnetic* field instead of electric field as, for example, in Fig. 102. These waves are known as *H* waves, and once again it is possible to have several orders of wave. Other possible waves are illustrated in Fig. 103.

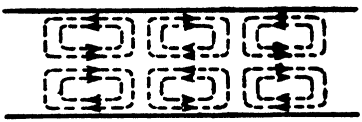

FIG. 102. MAGNETIC WAVE

Orders of Wave.

The explanations so far given have been of a physical character to enable the reader to visualize the manner by which waves can be propagated through metallic tubes. The full understanding of this form of transmission unfortunately involves a considerable knowledge of mathematics beyond the scope of this work, and it will be necessary in much of what follows to state the facts without proof. It is, however, helpful to have some understanding of the manner in which

different orders of waves can be set up in guides, and to know something of the nomenclature by which these different types of wave are distinguished.

We have seen that in general there are two types of wave known as the E and H types respectively, according to whether the axial component of the field in the direction of propagation is electric or magnetic. Each of these types can be further defined by two suffixes m and n. The first of these indicates the number of configurations within the tube. In the simplest type of wave there is only one such pattern. The wave of Fig. 100, for example, has one set of electric fields within the tube, and this wave is therefore known as an E_0 wave. In Fig. 101, on the other hand, we have two such patterns, each occupying one-half of the tube and this is known as an E_1 wave. With an E_2 wave we have four patterns each occupying a quadrant and so on.

With each of these types of wave, however, it is possible for the field distribution to have a number of harmonics, in the same way as a dipole aerial can oscillate at its fundamental or at a harmonic. The second suffix, therefore, indicates the number of nodes in the field. Thus as E_0 wave is strictly an E_{01} wave, while the form shown in Fig. 101 is an E_{11} wave.

A similar nomenclature is used for the H waves, dealing this time with magnetic field distribution instead of electric field. The diagrams shown in Fig. 103 illustrate some of the simpler types of wave in order to clarify this nomenclature.

The first wave is the simple E_{01} wave which we have just discussed. The electric fields are illustrated by full lines, and the magnetic fields by dotted lines. There will, therefore, be a series of circular bands of magnetic field coincident with the axial electric field as shown in the right-hand top diagram, where the clear circles represent the magnetic lines of force passing down through the paper, and the black circles represent the lines coming back again through the paper.

The second diagram shows an E_{02} wave. Here we have the same general distribution as before, but a harmonic distribution so that the electric field, while still radial, changes its sign in coming from the centre to the outside. There are two belts of magnetic field, an inner and an outer, both in opposite directions.

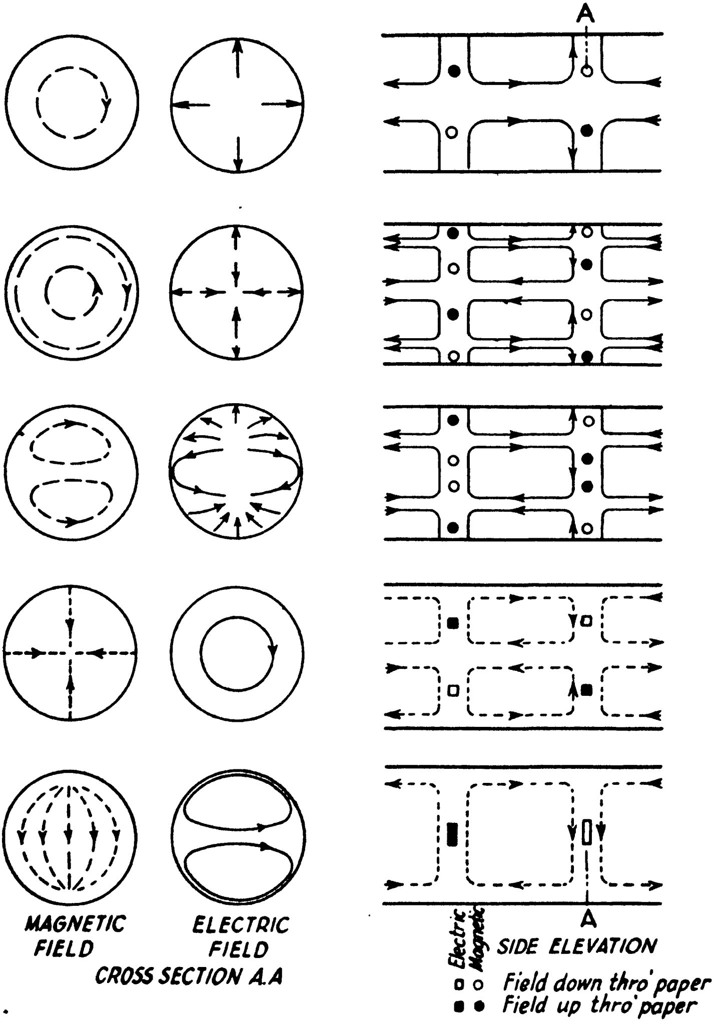

FIG. 103. ILLUSTRATING DIFFERENT TYPES OF WAVE. THEY ARE E_{01}, E_{02}, E_{11}, H_{01}, AND H_{11} RESPECTIVELY

The third diagram represents an E_{11} wave. It is as if the tube were divided across its horizontal diameter, and in both top and bottom halves we have an E_{01} wave. The two waves join up in the central region. The magnetic fields are distorted into a D shape formation.

The fourth diagram shows an H_{01} wave. This is similar to the E_{01} wave of the top diagram with the electric and magnetic fields interchanged, but whereas in all the E wave arrangements the electric field when reaching the metal guide completes itself within the guide itself, the magnetic field is entirely contained within the guide so that the complete loop of field appears.

The fifth diagram represents the H_{11} wave which is similar to the E_{11} wave of the third diagram, although it does not appear so at first. The electric field is split about a diameter, completing itself within the metal of the guide itself as already explained. The magnetic field will then take a form similar to the central portion of the electric field in the E_{11} diagram. The direction of the electric field in the top half of the guide will produce a magnetic field running down through the paper, while that in the bottom half will produce a field in the opposite direction. These magnetic fields will combine to form closed loops as shown in the right-hand diagram, and this produces a very simple type of wave which is used to a considerable extent in this technique.

It should be noted that the two electric fields in the central region are in the same direction. Hence, since the completion of the fields flows within the metal of the guide, the electric field inside the guide is simply a single transverse field running right across the guide. Hence the H_{01} wave gives the simplest field configuration of any which accounts for its popularity.

It is worth noting that the attenuation of the waves in a guide is always best with the lowest order of wave which is a further reason for the use of the simpler types of wave.

Rectangular Guides.

It is not necessary for the guides to be circular in shape. Rectangular guides may be used and often are employed, particularly in feeding horns, in order to obtain radiating

properties. The general mechanism of propagation in a rectangular guide is similar to that with a circular guide, but with the slight added complexity that it is possible for both the dimensions of the cross-section to have various forms of field distribution. Thus we can have a linear distribution of the field across the guide, and a sinusoidal distribution from top to bottom. We can have a sinusoidal distribution in both directions, or we can have harmonic distribution as in the case of circular guides.

We shall discuss the question of radiation from horns later, and the consideration of rectangular guides is best left until then to avoid confusion.

Critical Wavelengths.

It is found that for each type of wave there is a critical wavelength above which no transmission occurs. This critical wavelength depends upon the dimensions of the guide, and also upon the dielectric constant of the insulating material within the guide, or, strictly speaking, upon the relative dielectric constant of the dielectric guide itself and of the surrounding medium.

This means that it is theoretically possible to transmit waves along a guide consisting of a cylinder or other shaped section of dielectric alone without a metal tube around it, but such guides are not of practical value partly owing to the greatly increased attenuation which results if any solid dielectric is used. For practical purposes the dielectric is air enclosed in a metal cylinder, and since the dielectric constant of air is unity, the critical wavelength is dependent solely upon the dimensions of the guide.

CRITICAL CONDITIONS FOR WAVES IN CYLINDRICAL GUIDES

Type of Wave	Critical Wavelength (cm.)
E_{01}	$1{\cdot}31d\sqrt{k}$
E_{11}	$0{\cdot}82d\sqrt{k}$
H_{01}	$0{\cdot}82d\sqrt{k}$
H_{11}	$1{\cdot}71d\sqrt{k}$

d = diameter of guide (cm.)
k = dielectric constant of material inside the guide

The table herewith gives the critical wavelengths for various types of wave in circular guides. It is worth noting that the critical wavelength is of the same order as the diameter of the tube, which illustrates why practical progress in wave guides did not take place until centimetre wavelengths could be generated with reasonable facility.

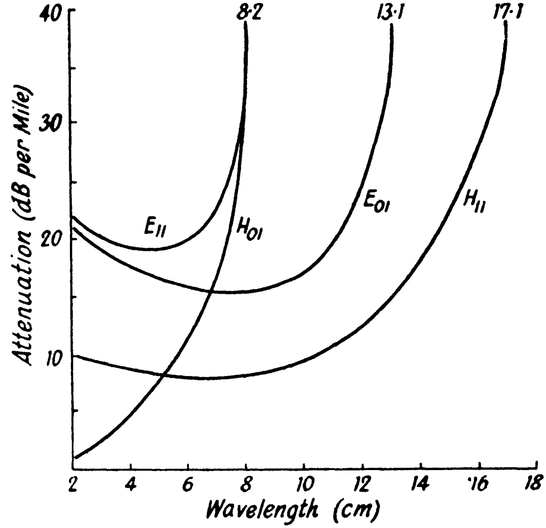

FIG. 104. ATTENUATION WITH DIFFERENT TYPES OF WAVE IN A 10-CM. TUBE

Attenuation.

The attenuation of the waves depends upon the wavelength relative to the critical wavelength for the particular conditions. At the critical wavelength, of course, the attenuation is infinite, i.e. the wave is not transmitted at all. Below the critical wavelength the attenuation drops rapidly and reaches a minimum at between one-third and two-thirds of the critical wavelength, depending on the type of wave. This minimum is fairly broad as illustrated in Fig. 104, and thereafter the attenuation rises slowly as the wavelength is progressively decreased.

An exception to this rule is with the H_{01} wave which has theoretically a continuously decreasing attenuation with decreasing wavelength. It has been found in practice, however, that the very slightest departure from true circularity of the guide introduces rapid attenuation, and the attenuation for an H_{01} wave in a practical tube, while rather better than that for an E_{11} wave, is nevertheless of the same form and passes through a minimum.

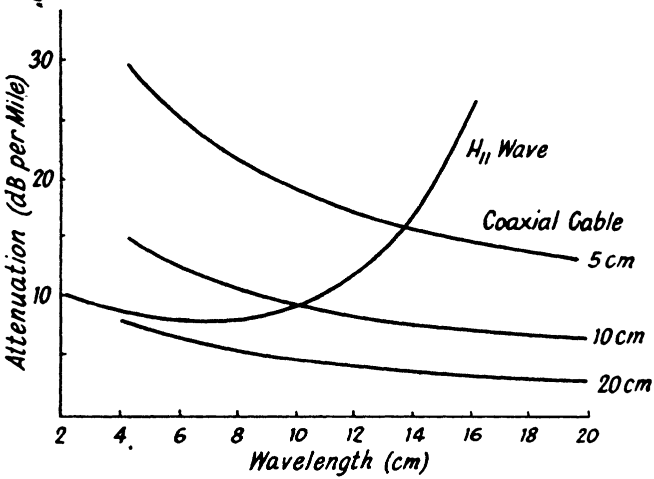

FIG. 105. COMPARISON OF 10-CM. WAVE GUIDE AND CO-AXIAL CABLE

For purposes of comparison Fig. 105 shows the relative attenuation of a 10-cm. wave guide transmitting an H_{11} wave and various sizes of coaxial cable. The improvement at low wavelengths is clearly seen.

Characteristic Impedance.

The absence of any current flow of conventional form causes the usual conception of characteristic impedance to break down. This impedance, it will be remembered, is the ratio of the voltage input to the cable divided by the current which it draws under suitable conditions of termination at the far end. We can, however, express the characteristic impedance

in terms of the power transmitted, and on this basis we find that the characteristic impedance is of the same order as for coaxial cable. Fig. 106 illustrates the characteristic impedances for a 10-cm. circular guide excited with different types of wave. Again, at the critical wavelength, the charac-

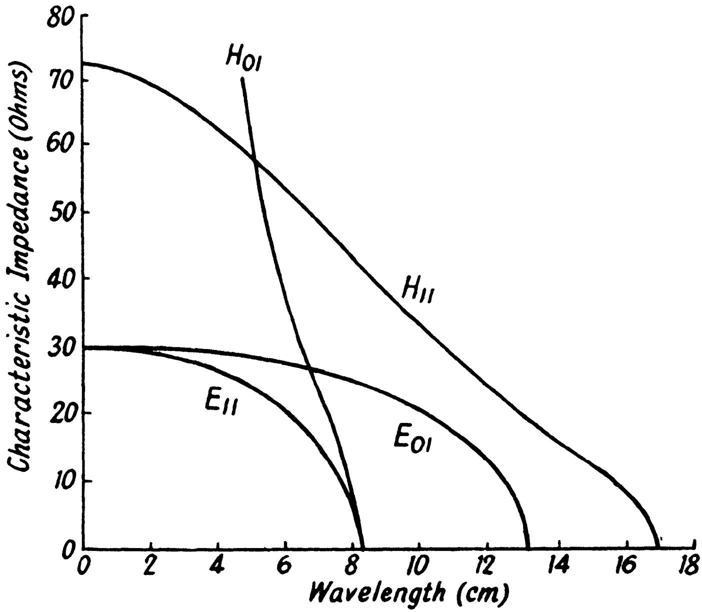

FIG. 106. CHARACTERISTIC IMPEDANCES OF 10-CM. GUIDE

teristic impedance is zero, but it rises rapidly and in the case of *E* wave rapidly approaches a constant value. With *H* waves the impedance increases more or less inversely as the wavelength.

Velocity of Propagation.

The velocity with which intelligence is transmitted through a wave guide is always less than with a wave travelling in free space, the actual velocity depending on the wavelength relative to the cut-off wavelength for that mode of vibration.

At cut-off the velocity of propagation is zero because no energy is transmitted at all. As the frequency is increased the velocity of propagation increases until at wavelengths well below the cut-off wavelength the velocity of propagation becomes very nearly that of light, i.e. the same as in free space.

Certain mathematical conceptions in the expressions dealing with wave guides result in velocities exceeding that of light, but these are in fact mathematical fictions and the velocity with which intelligence can be transmitted is always less than that of light, as in fact it must be.

Launching the Waves.

The waves are first set up or *launched* in the guide by short rods or aerials arranged along the direction of the electric component of the wave required. Thus, in Fig. 100 where we launched an E_0 wave, the aerial was arranged along the axis of the tube. This particular aerial would be fed by a parallel wire or coaxial feeder coming in through the side of the guide at right angles to the plane of the paper.

A more usual arrangement is to feed the energy into the guide from a coaxial cable, which may or may not be of the same diameter as the guide itself. In this case it is simply necessary to extend the central conductor of the coaxial cable along the guide for a short distance, as shown in Fig. 107 (*a*). Similarly, if we wish to launch an E_1 wave, we can do this by having two dipole aerials as in Fig. 101, or by having two rods protruding into the inside of the guide as shown in Fig. 107 (*b*). In both cases it is necessary that the phase of the voltage on the two aerials or rods is in opposition, which is conveniently done by introducing an additional half wavelength of cable between one aerial and the other.

With an H_1 wave there is, as we have seen, a single electric field in the centre of the tube, and this may therefore be introduced by a single rod placed along a diameter of the tube as shown in Fig. 107 (*c*). With an H_0 wave we have two such fields, and these must be generated by two rods each extending across approximately half the diameter of the tube as in Fig. 107 (*d*).

There are various other ways of launching waves in guides which cannot be discussed here. Further information on this

point will be found in the paper by Kemp, referred to at the end of the chapter. This same article also contains information as to how one type of wave may be transformed into another by interposing barriers or gratings within the guide. This is sometimes useful, for it may be that more than one type of wave is present at the same time and the undesired type of wave may be trapped by the interposition of suitable barriers designed to coincide with the configuration of the electric field, but of such a nature that they do not seriously interfere with

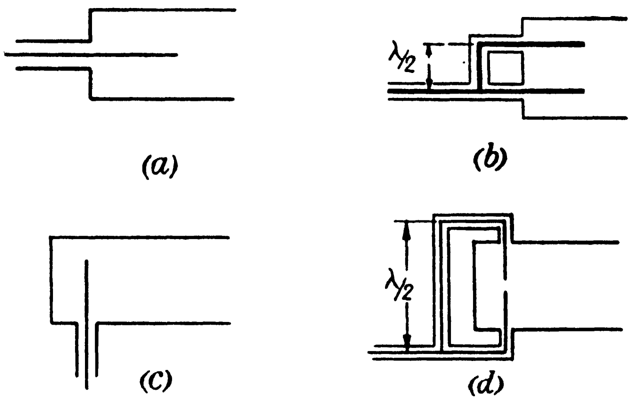

FIG. 107. METHODS OF LAUNCHING WAVES IN GUIDES

the substantially different field distribution for the wanted wave.

Reception of Waves.

The arrangement used for launching waves may be used equally for taking the waves out of the guide at the far end, and transferring the energy to a parallel wire feeder or a coaxial cable. Alternatively, we can arrange gratings so disposed that they lie along the configuration of the electric field for the particular type of wave. If, in addition, these gratings contain a non-linear element such as a crystal detector, the current which will flow will be unidirectional and will therefore be able to operate an indicating device. A series of possible arrangements are shown in Fig. 108. The first is for detecting

an E_{01} wave and contains a number of wires arranged along the lines of electric force, each containing a crystal. The second figure is for an H_{01} wave where the electric field is circular. The grating is thus in the form of a nearly closed loop. If the length of the wire in this loop is comparable with the wavelength several crystals may be used in series, suitably spaced as shown. The third figure shows a very simple detecting grating for an H_{11} wave where, as we have seen, the electric field is straight across the tube. The small barriers

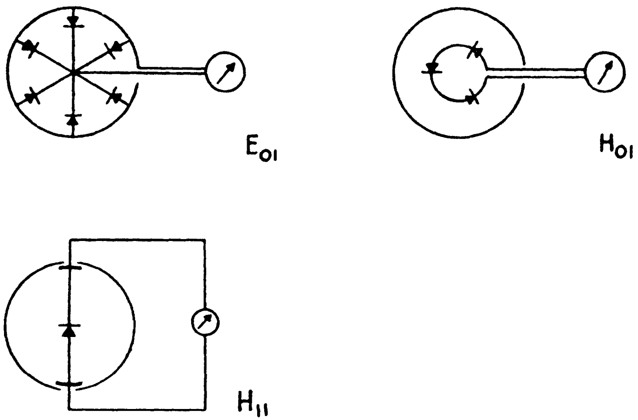

FIG. 108. ARRANGEMENTS FOR DETECTING WAVES IN GUIDES

at each end of the wire form blocking condensers and prevent the field from escaping.

Resonance Cavities.

We have already referred to the possibility of cutting off or reflecting the waves in any unwanted direction by means of a complete barrier across the guide. Thus, in Fig. 100, we could interpose at the left-hand side of the guide a barrier situated a correct multiple of half a wavelength away from the aerial, in such a manner that the waves radiated to the left were reflected and arrived back at the starting point in such a phase as to add to the waves travelling towards the right. The existence of longitudinal waves within the guide is

in fact exactly analogous to the existence of longitudinal air vibrations in an organ pipe. In order to adjust the reflections correctly, this barrier is often made in the form of a movable piston, the position of which can be altered until the maximum effect is obtained.

A similar arrangement may be used at the receiving end, and the reader will appreciate that there are many other possibilities whereby the length of the tube at a particular point can be adjusted to produce a correct phasing of the vibration. One immediate application of this idea is the closure of the guide at both ends, one end by means of an adjustable piston, in order to form an enclosed cavity in which the waves will resonate and produce standing waves in the same way as in a closed organ pipe. Such resonating chambers may be used in association with an ordinary transmission guide for the purpose of accentuating the effect at any particular point.

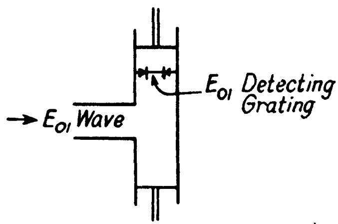

FIG. 109. RESONANCE CAVITY

Fig. 109, for example, shows an ordinary wave guide terminated in a second guide at right angles. The electric fields transmitted along the guide creep round the corners into the resonating cavity, where they set up standing waves which may be adjusted to their maximum intensity by altering the length of the resonance chamber by means of the pistons shown. A detecting grating is located at a point of maximum field intensity, and the result so produced will be many times greater than that obtained if the detecting grating were located in the guide itself. The effect is in fact exactly similar to the resonant action of an ordinary tuned circuit. A similar arrangement may be used for introducing a heterodyne, an oscillator of the required frequency being located in a second guide which also feeds into the resonating cavity.

The fact that the form of detecting grating is different according to the type of wave being received makes it possible for two waves to be transmitted along the same guide, and for these to be received at the far end quite separately and

distinctly by using two detecting gratings placed a little way apart.

There are various other possibilities, too numerous to be discussed here. Mention may be made of the possibility of building an oscillator into the guide. This is an arrangement which is not only useful for the generation of waves, but permits the construction of receivers in which the valve is used not as an oscillator but as an amplifier, receiving energy at a low level and feeding it to a valve which, in association with a series of tuned cavities, produces a considerably higher level of energy, which is then launched on to a fresh guide.

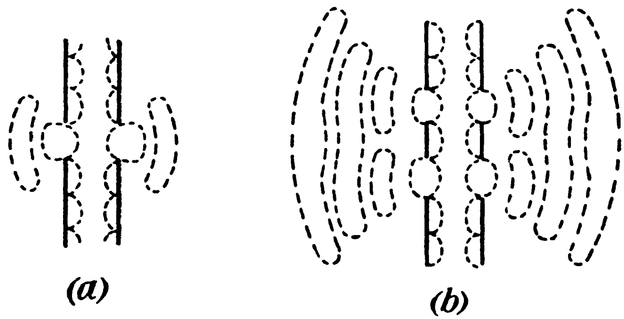

FIG. 110. ILLUSTRATING RADIATION FROM A WAVE-GUIDE

Details of these other applications will be found in a paper by John Kemp entitled "Wave Guides in Electrical Communications," *Journal I.E.E.*, Part III, No. 11 (September, 1943).

This paper also contains an extensive bibliography, citing some fifty-two papers and books to which the reader may refer for further information.

Radiation from Guides.

We saw in Fig. 100 that a wave could be launched in a guide by locating the aerial within the guide, so that the electric fields instead of being able to spread out were confined within the guide. It is clear that if we leave a gap in the guide at a suitable point these fields will emerge from their constraint, and will again begin to radiate. A simple example of this is illustrated in Fig. 110, where a gap has been left in a guide

carrying an E_{01} wave and the manner in which the electric field will emerge is quite clear. If we leave two gaps a suitable distance apart as shown in Fig. 110 (*b*), we have two emergent loops of field which will join up and form one field of grater amplitude. If, moreover, the end of the guide is terminated as a resonating cavity, the field will be greatly amplified, and hence if gaps are provided in this resonating cavity the radiation will be considerably enhanced.

If the width of the gap is restricted the energy is radiated in the form of a relatively narrow beam, and this radiation may be assisted by the provision of conical flanges. Fig. 111 shows such an arrangement with the end of the guide proper terminated by a movable piston, in order to obtain the desired resonant effect, and there are numerous other possibilities which need not be discussed in detail.

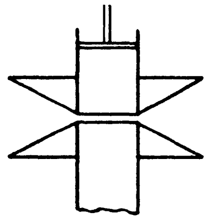

FIG. 111. FLANGED RADIATOR

Rectangular Guides as Radiators.

When the guide is to be used for radiation, however, it is found that the use of a rectangular guide has much to commend it. As already explained, waves can be set up in rectangular guides in the same manner as in circular guides, though the type of wave is, in general, of a rather more complex order because of the possibility of having various distributions of field along *both* dimensions of the rectangle. For the present purposes we shall confine our attention to the simplest type of wave only, which is known as the H_{01} wave. The nomenclature for waves in rectangular guides is not the same as for circular guides, though it follows a similarly logical system. The essential difference is that it is possible to have waves in which the field distribution along one of the dimensions of the rectangle is linear, as a result of which the starting point is, as it were, one degree lower than with circular guides. The H_{01} wave referred to is one in which the magnetic field is distributed across one dimension of the guide, and is of the same intensity all the way across so that its distribution is linear. The electric field is at right angles to the magnetic field, and again flows straight across the guides

but its distribution is sinusoidal. (A wave having linear distribution in both directions does not exist.*)

As in the case of circular guides, there is a critical frequency below which waves are not propagated, and in the case of a simple H_{01} guide just mentioned this critical frequency is twice the width of the tube and is independent of the height.

The distribution of field in an H_{01} wave in a rectangular tube is as shown in Fig. 112, and its convenience as a radiator will at once be apparent in view of the fact that the electric

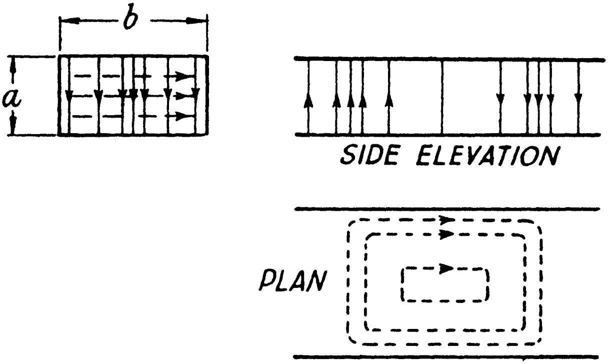

FIG. 112. FIELD CONFIGURATIONS WITH H_{01} WAVE IN RECTANGULAR GUIDE

field lies straight across the guide from top to bottom, this travelling field being accompanied by single configuration of magnetic field. If such a guide is left open at the far end the electric field will simply continue to travel out into space.

There will also be an appreciable directive effect, the waves being radiated within a fairly narrow angle depending upon the dimensions of the guide relative to the wavelength.

The ratio of the length of the side (a) or (b) relative to the wavelength is called the *aperture*, and it is found that for a given aperture the beam is sharper in the vertical plane than in the horizontal plane in the ratio of 2 to 3. Thus with a

* Both E and H waves can exist in rectangular guides and both types may be of the type E_{mn} or H_{mn} where the suffixes mn determine the number of half sinusoids in the distribution of field intensity along the two sides of the rectangle respectively.

square section tube the radiation will only spread vertically to two-thirds of the horizontal spread.

As in general it is desirable to restrict the spreading of the beam in the horizontal plane to a greater extent than the spreading in the vertical plane, the form of guide used in practice is appreciably wider than it is high (remembering that the wider the guide the greater the aperture, and hence the *narrower* the beam). Fig. 113 shows the relation between

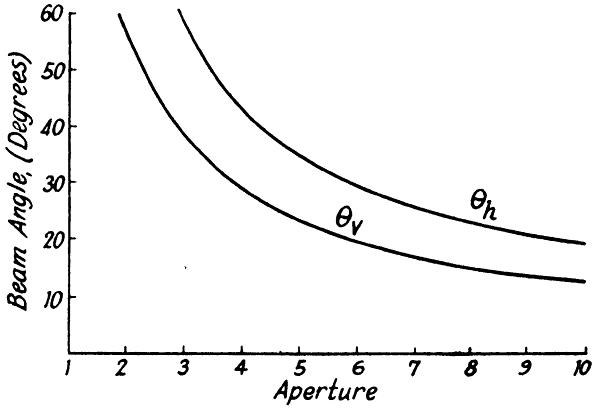

FIG. 113. RELATION BETWEEN APERTURE AND BEAM ANGLE

the aperture of the horn and the angle of the radiated beam. There are two curves because of the difference between the vertical and horizontal aperture effect already mentioned, and it will be seen that with apertures between 5 and 10 a very narrow beam may be produced.

Use of Horns.

Still further improvement in the radiation may be obtained by flaring the mouth of the tube to form a horn. This may be done in a horizontal dimension only or in both directions according to the requirements. As might be expected, there is an optimum angle of flare which usually lies between 30° and 50°, while there is also a minimum length of horn. The horn acts as a resonating cavity, and thus amplifies the

strength of the fields, but if the horn is less than the critical length the greater part of the energy is confined within the horn and does not radiate. Beyond the critical length, however, an increasing amount of energy is radiated, and because of the resonance effect just mentioned it will be seen that the radiated field is greater than it would be from a simple open-ended tube.

The horn also increases the directional effect. We have seen that with a plain tube the beam angle becomes progressively less as we increase the dimensions of the tube, but this may require an inconveniently large construction. A correctly designed horn produces the same sharpness of beam as a simple guide of the same dimensions as its outer extremities.

The theory has been developed by various investigators, and it shows that for wavelengths below 50 cm. the electromagnetic horn is superior to all other types of radiator.

For further details on wave guides and kindred subjects the reader is referred to—

"Transmission of Electromagnetic Waves in Hollow Tubes of Metal." Barrow. *Proc. I.R.E.*, Vol. 24, p. 1298 (1936).

"Some Fundamental Experiments with Wave Guides." Southworth. *Proc. I.R.E.*, Vol. 25, p. 807 (1937).

"Theory of the Electromagnetic Horn." Barrow and Chu. *Proc. I.R.E.*, Vol. 27, p. 51 (1939).

"Wave Guides in Electrical Communications." Kemp. *Journal I.E.E.*, Part III, No. 11 (1943).

CHAPTER XIII

VELOCITY MODULATED OSCILLATORS

THE valves used for normal short-wave operation embody an electron stream of constant velocity, but varying strength. To reduce the amplitude some of the electrons in the stream must be cut off, i.e. their velocity must be reduced substantially to zero. At low frequencies the time required for this to happen is negligible in comparison with the oscillation period.

At high frequencies this is no longer true. The time taken by the electron to change its velocity becomes comparable with the oscillation period, and in such circumstances the velocity of the electron stream can no longer be considered as constant, even though its average value may be. In a B–K oscillator, for example, the electrons are continually changing their velocity as they oscillate to and fro past the grid.

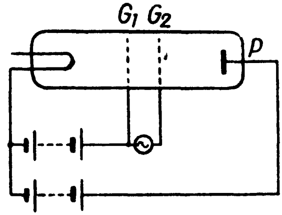

FIG. 114. VELOCITY MODULATED TUBE

This led various investigators to consider the possibility of controlling the electron stream by varying its velocity instead of its amplitude.

Such an arrangement is illustrated in Fig. 114. The grid G_1 is at a high potential and produces an electron stream from the cathode. G_2 is varied above and below the potential of G_1 by some small fraction, and this causes the electrons to be alternately accelerated and decelerated.

In the space between G_2 and the plate P therefore we have a velocity modulated stream. There are three ways in which we can utilize this.

Retarding Field Conversion.

If we reduce the voltage on the plate P to a low value we can arrange that electrons travelling with normal or reduced

velocity are turned back, so that only those electrons which have greater velocity than normal manage to reach the plate. Hence we obtain a plate current depending upon the modulation, thus converting the velocity modulation into an amplitude change.

Such an arrangement has been used with success. It is necessary to provide an additional electrode to collect the electrons which are turned back from the plate, while it is further essential for reasonable efficiency that the external impedance in the plate circuit shall be high at the oscillation frequency, and this is not easy to achieve. The use of resonant

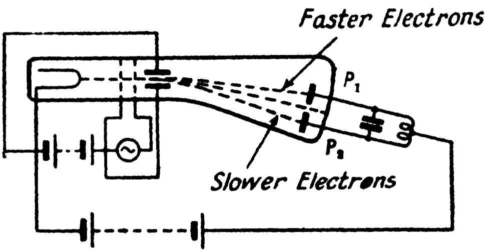

FIG. 115. DEFLECTION OSCILLATOR

cavities similar to those discussed in the last chapter enables improved results to be obtained, but if this technique is used it is found that better methods can be applied, as we shall see, and the retarding field generator is little used.

Deflection Generators.

A second method is to deflect the electron stream as in a cathode ray tube, when the more rapidly moving electrons are deflected less than the normal or slower moving ones. By using a split collector as shown in Fig. 115, all the accelerated electrons will arrive at plate P_1, while the decelerated electrons, being deflected farther, arrive at P_2, and thus again we convert the frequency modulation into a change of amplitude.

This arrangement, however, suffers from the disadvantage that a high external impedance is essential, while it is also insensitive to small changes of velocity, since there must be a gap between P_1 and P_2, and unless the velocity change is

sufficient the beam will not be deflected enough to reach either plate. Similarly, at high levels of input, the beam is completely separated into two components, and no further increase in output results, so that the device saturates.

Drift Tube Converters.

The third and most successful method is to allow the electron stream to convert itself into an amplitude modulated stream by simply allowing it to continue on its way. The faster moving electrons then overtake the slower moving ones, and since the

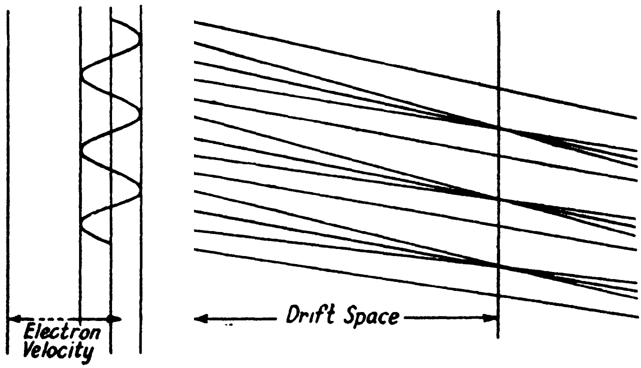

FIG. 116. PRINCIPLE OF DRIFT TUBE

velocity has been modulated periodically, there will be points along the *drift tube* (the space in which the electrons are allowed to proceed uninterrupted) where there will be accumulations or bunches of electrons.

Fig. 116 illustrates this graphically. The various lines represent electrons entering the drift tube at regular intervals. The slope of the lines represents the velocity which will be seen to increase and decrease rhythmically. It will be seen that the majority of the electrons arrive at end of the tube together.

The process may be illustrated by a simple analogy. Suppose we dispatch a series of trains at five-minute intervals travelling at 30 m.p.h. If some time afterwards we dispatch another train on a parallel line running at 60 m.p.h. it will obviously

catch up with each of the slower-moving trains in turn, and it is not difficult to see that if we had a series of the faster trains dispatched at five-minute intervals there would be certain points along the track at which the overtaking took place, and the inhabitants of any one of these places would always see two trains going through together, followed by an interval with no trains at all, so that these would obviously be good spots at which to indulge in the pastime of watching the trains go by.

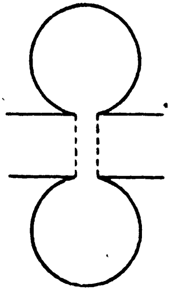

FIG. 117
RHUMBATRON

Rhumbatrons.

The electrons will take a finite time in passing between the grids G_1 and G_2 of Fig. 114. Clearly this time must not be more than the period of one half cycle of the modulating frequency. Otherwise an electron which had been accelerated immediately after passing G_1 would be decelerated again before it had got clear of G_2.

This requirement leads to the use of a combined oscillator and modulating grid structure of the form shown in Fig. 117, to which the name *rhumbatron* was given by the brothers R. H. and S. F. Varian, who first evolved a practical drift tube type of oscillator. Fig. 118 illustrates the development

FIG. 118. DEVELOPMENT OF RHUMBATRON FROM *L–C* CIRCUIT

of this form of oscillator from the more conventional form of circuit.

On the left we have a condenser with two circular plates joined by a single turn inductance. Such a circuit is limited in frequency by the inductance of the loop, while it is also of poor Q because the greater part of the energy is radiated (whereas a high Q implies that the greater part of the energy is *stored*). The second figure shows two loops in parallel, which reduces the effective inductance and also reduces the

radiation since the fields from the two loops tend to cancel out. Hence both f and Q have been increased.

The more loops we add the greater the improvement and the ultimate configuration is a complete toroidal sheet of copper with the condenser in the centre, which is the shape of Fig. 117. Such a device is an oscillatory circuit of very high Q since there is no radiation—the Q without external loading being of the order of 10 000. Even when appreciable power is taken from the circuit we can still maintain a Q of the order

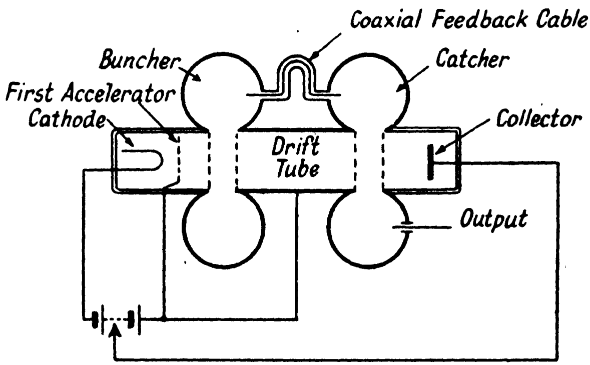

FIG. 119. KLYSTRON OSCILLATOR

of 1 000, which is far better than is attainable by normal methods.

The Klystron.

The brothers Varian, already mentioned, used such a device in the development of a practical oscillator. The centre section was made with a grid structure, and replaced the two grids G_1 and G_2 of Fig. 114. Oscillation is set up in this rhumbatron, which is called the *buncher*, by introducing a short rod or wire along a line of electric force, similar to the manner in which E waves are set up in a wave guide. This is followed by a drift tube and a second rhumbatron known as the *catcher* where the arrival of the bunched electrons excites oscillations. To render the whole system self-maintaining a collector in the catcher

extracts some of the energy, and feeds it back to the buncher, thus maintaining the oscillations therein.

When the device is working there is a strong field across the grids of the catcher, and this is in such a phase as to retard the electrons and thus extract most of energy from them. They pass through the catcher at quite low velocity, and are collected by a final electrode at low potential. Thus the energy in the external circuit is quite small, most of the energy having been absorbed by the catcher, so that the arrangement is highly efficient. Powers of 300 watts of 10 cm. are readily generated by this type of oscillator.

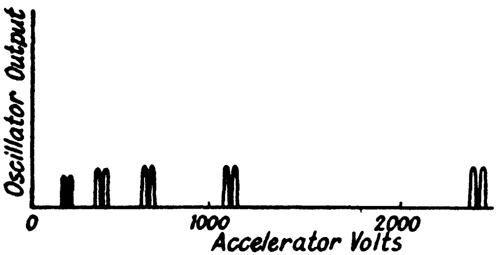

FIG. 120. OSCILLATION CHARACTERISTIC OF KLYSTRON

Operating Potential.

The velocity of the electrons must be such that the transit time across the grids of the buncher and also along the drift tube is such as to provide the correct phase relations. The velocity is controlled by the potential on the first accelerator, and hence there is a series of correct operating potentials at which oscillation is possible. At all other potentials the device does not operate. Moreover, since catcher and buncher are tightly coupled the system has two adjacent frequencies (as with any coupled circuit system), each with its own range of operating potential. The oscillation characteristic is thus as shown in Fig. 120.

The Klystron may be used as an amplifier as well as an oscillator by feeding energy to the buncher, and taking amplified energy from the catcher. It is not a linear amplifier except at low inputs. With larger inputs the drift tube

bunches practically all the electrons so that further increase is not possible. This effect has been used to provide a frequency multiplying arrangement, the catcher being tuned to a multiple of the buncher frequency.

The fixed frequency of the Klystron is perhaps its only serious disadvantage. This is partially offset by making the device mainly of metal with a glass portion containing the cathode, and first acceleration sealed on to the input end. The rhumbatrons may then be made with corrugated sides

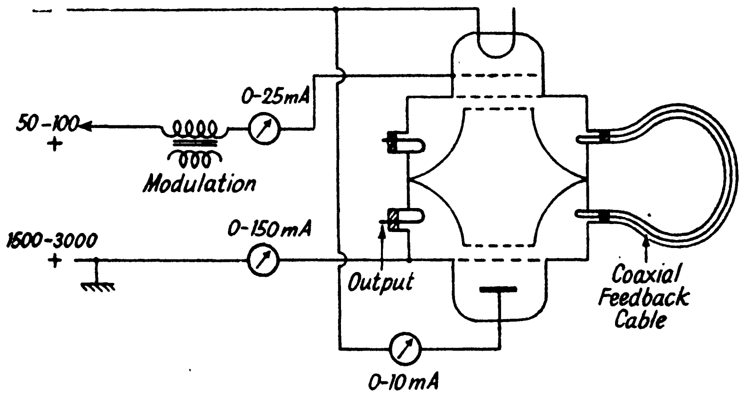

FIG. 121. MODIFIED FORM OF KLYSTRON

enabling their shape, and hence their resonant frequency, to be changed over a small range. The shape of the cavities is different from the theoretical toroid, as indicated in Fig. 121, which also shows some of the auxiliary circuits.

Reflex Klystron Oscillator.

A modification of these arrangements uses only one rhumbatron. The distance from the exit grid to the final anode and the potential on the said anode are then so arranged that the velocity modulated electrons are turned back and re-enter the exit grid. As they have bunched in the process their energy is rapidly extracted and self-oscillation builds up. The action is somewhat similar to that of the *B–K* oscillator. A diagram of the circuit is shown in Fig. 122.

Inductive Output Amplifier.

The latest development in this class of oscillator constitutes a break in certain important particulars from the Klystron technique, and results in a rather more flexible arrangement.

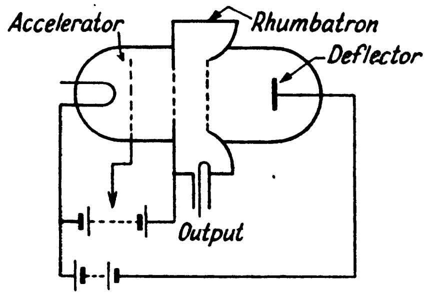

FIG. 122. REFLEX KLYSTRON

This system uses what is called an inductive output tube, and is illustrated in Fig. 123. A simple grid is used both as first accelerator and modulator. The arrangement is more an amplitude rather than a velocity modulator, the effect being to generate bunches of electrons which are then accelerated down the tube by additional accelerators.

Outside the tube is a cylindrical resonator with an annular

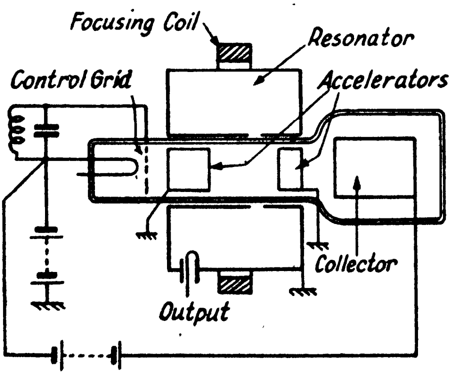

FIG. 123. INDUCTIVE OUTPUT OSCILLATOR

gap, across which strong electric fields appear when the resonator is oscillating. If the phase is correctly adjusted these fields will be of maximum intensity when a bunch of electrons is passing, and if they are in such a direction as to retard the electrons they will extract energy from them. In practice the energy can be almost wholly extracted, leaving only a small residual to be drained off by the final ánode.

A further improvement results from the application of an axial magnetic field which prevents the electrons from scattering. It imparts a spiral motion tending to concentrate the electrons in a beam as in a magnetically focused cathode-ray tube, and the arrangement as a whole forms a highly efficient mechanism for amplifying or generating oscillations between 100 and 1 000 Mc/s.

The fact that the resonator is external to the tube has obvious advantages, since the tube is separated from the associated circuits as with the more conventional forms of valve. There is, of course, still the need to adjust the operating potentials within critical limits to maintain the correct phasing.

An excellent review of the whole subject will be found in an article by Sarbacher and Edson entitled "Tubes Employing Velocity Modulation," *Proc. I.R.E.*, August, 1943, while further details may be obtained from the following papers—

"Velocity-modulated Tubes." Hahn and Metcalf. *Proc. I.R.E.*, Vol. 27, February, 1939.

"A High-frequency Oscillator and Amplifier." R. H. and S. F. Varian. *Journ. App. Phys.*, Vol. 10, May, 1939.

"Resonators Suitable for Klystron Oscillators." Hansen and Richtmeyer. *Journ. App. Phys.*, Vol. 10, March, 1939.

"Theory of Klystron Oscillator." Webster. *Journ. App. Phys.*, Vol. 10, December, 1939.

"Beyond the Ultra-short Waves." Southworth. *Proc. I.R.E.*, September, 1943.

INDEX